



Michigan Technological University
Create the Future Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's
Reports - Open

Dissertations, Master's Theses and Master's
Reports

2007

Analysis of atmospheric transport and the effects of seasonal and interannual transport variability on measurements made at the Pico Mountain observatory

Jessica Marie Strane
Michigan Technological University

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Civil and Environmental Engineering Commons](#)

Copyright 2007 Jessica Marie Strane

Recommended Citation

Strane, Jessica Marie, "Analysis of atmospheric transport and the effects of seasonal and interannual transport variability on measurements made at the Pico Mountain observatory", Master's Thesis, Michigan Technological University, 2007.
<https://digitalcommons.mtu.edu/etds/272>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Civil and Environmental Engineering Commons](#)

**An Analysis of Atmospheric Transport and the Effects of
Seasonal and Interannual Transport Variability on
Measurements Made at the Pico Mountain Observatory**

By

Jessica Marie Strane

A THESIS

Submitted in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE in ENVIRONMENTAL ENGINEERING

MICHIGAN TECHNOLOGICAL UNIVERSITY

2007

Copyright © 2007 Jessica Strane

This thesis, “An Analysis of Atmospheric Transport and the Effects of Seasonal and Interannual Transport Variability on Measurements Made at the Pico Mountain Observatory,” is hereby approved in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE in ENVIRONMENTAL ENGINEERING.

Department of Civil and Environmental Engineering

Dr. Richard E. Honrath, Thesis Advisor

Dr. Neil Hutzler, Chair, Department of Civil and Environmental Engineering

Date

Abstract

A non-hierarchical K-means algorithm is used to cluster 47 years (1960–2006) of 10-day HYSPLIT backward trajectories to the Pico Mountain (PM) observatory on a seasonal basis. The resulting cluster centers identify the major transport pathways and collectively comprise a long-term climatology of transport to the observatory. The transport climatology improves our ability to interpret the observations made there and our understanding of pollution source regions to the station and the central North Atlantic region. I determine which pathways dominate transport to the observatory and examine the impacts of these transport patterns on the O_3 , NO_y , NO_x , and CO measurements made there during 2001–2006. Transport from the U.S., Canada, and the Atlantic most frequently reaches the station, but Europe, east Africa, and the Pacific can also contribute significantly depending on the season. Transport from Canada was correlated with the North Atlantic Oscillation (NAO) in spring and winter, and transport from the Pacific was uncorrelated with the NAO. The highest CO and O_3 are observed during spring. Summer is also characterized by high CO and O_3 and the highest NO_y and NO_x of any season. Previous studies at the station attributed the summer time high CO and O_3 to transport of boreal wildfire emissions (for 2002–2004), and boreal fires continued to affect the station during 2005 and 2006. The particle dispersion model FLEXPART was used to calculate anthropogenic and biomass-burning CO tracer values at the station in an attempt to identify the regions responsible for the high CO and O_3 observations during spring and biomass-burning impacts in summer.

Acknowledgements

I would like to thank the Environmental Engineering Department of Michigan Tech for the financial support and for the excellent engineering education I have received, my advisor Richard Honrath, for taking me on as an undergraduate research assistant and allowing me to continue with the group as a Master's student, and the rest of the Pico group for their advice and support. I am very grateful to my friend and fellow group member Chris Owen who, in addition to patiently teaching me the major programs used for this research, was always willing to answer my questions and prepare me for Richard's. My parents, if they feel pride in my accomplishments, should also be proud of themselves. Their trust and support has made me trust in my own abilities. Finally, I thank Kyle for listening with a sympathetic ear and for being so patient.

Contents

List of Figures	v
List of Tables	viii
1 Introduction	1
2 Methods	4
2.1 Trajectories and the Clustering Algorithm	4
2.2 Measurement Techniques	8
2.3 Upslope Flow Effects at Pico Mountain	8
2.4 FLEXPART Simulations	9
2.5 A Note on Chemistry	10
3 Results and Discussion	11
3.1 Transport Pathways to the Pico Mountain Observatory	11
3.2 Seasonal and Interannual Variability of Transport to the Pico Mountain Observatory	12
3.3 Effects of Transport Pathways on the Chemical Observations Made at the Pico Mountain Observatory	17
3.4 Comparison to Previous Studies at the Pico Mountain Observatory .	25
3.5 FLEXPART CO Tracer Analysis for 2003–2005 Spring	30

4	Summary and Conclusions	37
5	Future Work	40
A	Additional Detail on Analyses to Determine the Final Clustering Method	42
A.1	Describing Clusters: Variance	42
A.2	Major Changes to the Owen 2003 Clustering Code	44
A.3	Seed Sensitivity Analysis: Methods and Results	45
A.4	Clustering Method Final Decision: Mixed Flow versus Non-divergent	50
B	1960–2006 Spring, Fall, and Winter Frequency Plots at the Pico Mountain Observatory	56
C	Spring and Winter Transport Correlations with the NAOI	60
D	Time Series of 1960–2006 Seasonal Total Transport Percents	63
E	Examining the Uncertainty of the 10-Day HYSPLIT Backward Trajectories	68
F	Analysis of Upslope-Flow Screening on the CO and O₃ Measurements	73
	References	78

List of Figures

3.1	1960–2006 Non-divergent Seasonal Cluster Centers	13
3.2	1960-2006 Summer Cluster Area-Normalized Frequency	14
3.3	Mean Spring Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006	19
3.4	Mean Summer Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006	21
3.5	Mean Fall Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006	22
3.6	Mean Winter Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006	24
3.7	2001-2006 Summer Non-divergent and Weighted Mixed-Flow Cluster Mean Comparison	29
A.1	1960-2006 Spring Cluster Centers Produced with Convergent Seeds	47
A.2	1960-2006 Spring Cluster Centers Produced with 2,000 Trajectory Limited Seeds	47
A.3	1960-2006 Spring Cluster Centers Produced with Random Seeds . .	48
A.4	1960-2006 Summer Cluster Centers Produced with Convergent Seeds	48
A.5	1960-2006 Summer Cluster Centers Produced with 2,000 Trajectory Limited Seeds	49
A.6	1960-2006 Summer Cluster Centers Produced with Random Seeds .	49

A.7	1960–2006 Mixed-Flow Seasonal Cluster Centers	52
A.8	1960–2006 Non-divergent Seasonal Cluster Centers	53
B.1	1960–2006 Spring Cluster Area-Normalized Frequency	57
B.2	1960–2006 Fall Cluster Area-Normalized Frequency	58
B.3	1960–2006 Winter Cluster Area-Normalized Frequency	59
C.1	Linear Regression of the 1960–2006 Spring Cluster Transport and the North Atlantic Oscillation Index (NAOI)	61
C.2	Linear Regression of the 1960–2006 Winter Cluster Transport and the North Atlantic Oscillation Index (NAOI)	62
D.1	Time Series of the Total Transport Percents for Each Year in the Spring Seasonal Cluster Center from 1960–2006	64
D.2	Time Series of the Total Transport Percents for Each Year in the Summer Seasonal Cluster Centers from 1960–2006	65
D.3	Time Series of the Total Transport Percents for Each Year in the Fall Seasonal Cluster Centers from 1960–2006	66
D.4	Time Series of the Total Transport Percents for Each Year in the Winter Seasonal Cluster Centers from 1960–2006	67
E.1	Linear Regression of 5-Day Non-divergent Transport Percents against 10-Day Non-divergent Transport Percents	70
E.2	Linear Regression of 5-Day Total Transport Percents against 10-Day Total Transport Percents	71
F.1	Comparison of Upslope Flow Screened and Unscreened Spring CO and O ₃ Means by Cluster Center	74
F.2	Comparison of Upslope Flow Screened and Unscreened Summer CO and O ₃ Means by Cluster Center	75

F.3	Comparison of Upslope Flow Screened and Unscreened Fall CO and O ₃ Means by Cluster Center	76
F.4	Comparison of Upslope Flow Screened and Unscreened Winter CO and O ₃ Means by Cluster Center	77

List of Tables

3.1	2001–2006 Percent Transport in the 1960–2006 Seasonal Cluster Centers	16
3.2	Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by Source for 2002, 2003, 2004, and 2005 Springs	33
3.3	Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by Source for 2002, 2003, 2004, and 2005 Summers	34
3.4	Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by Source for 2002, 2003, and 2004 Falls	35
3.5	Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by Source for 2002, 2003, and 2004 Winters	36
A.1	1960-2006 Seasonal Seed Sensitivity Analysis of Cluster Variance . . .	46
A.2	Variance Comparison for the 1960–2006 Winter Mixed-Flow and Non- divergent Cluster Centers	54

Chapter 1

Introduction

The Pico Mountain (PM) observatory (formerly the PICO-NARE station) is located at 2225 m asl in the caldera of Pico Mountain in the Azores Islands in the central North Atlantic. The station is isolated from nearby pollution sources and ideally situated to intercept pollution plumes that affect air quality over the larger North Atlantic region. In operation since the summer of 2001, the PM observatory frequently samples free tropospheric air containing anthropogenic emissions and boreal outflow from North America and clean marine flow from the Atlantic (both in individual pollution plumes and mixed together), as well as less frequent flow from Europe, the Pacific, Asia, and Africa. Michigan Technological University (MTU) oversees measurements of ozone (O_3), carbon monoxide (CO), reactive odd nitrogen (NO_y), and nitrogen oxides (NO_x); the University of the Azores and more recently the University of Colorado collaborate with MTU to measure black carbon (BC) and non-methane hydrocarbons (NMHCs), respectively. In 2006, the station was transferred to the University of the Azores with the aim of establishment as a permanent observatory. After being shut down during 2007, the PM observatory will continue operation for the 2008 and 2009 summers with the addition of a peroxyacetyl nitrate instrument, run by the National Oceanic and Atmospheric Administration.

The longer-term measurement record and unique location make the Pico Mountain observatory an important site for studying atmospheric chemistry and transport in the North Atlantic region; understanding the major transport pathways and source regions is vital for better interpretation of the measurements made there. The analysis presented in this paper uses a trajectory clustering technique in combination with measurements from the station to identify transport pathways and to quantify the transport from potential source regions and their impacts on observations at the station.

Cluster analysis is a statistical technique used to illustrate the major patterns in large data sets, and the clustering of atmospheric trajectories has been used since the late 1980's to study atmospheric transport [*Harris and Kahl*, 1990]. The clustering process in combination with a large trajectory data set reduces the effect of the error inherent in the individual trajectories, which is on average 20% of the distance traveled [*Stohl*, 1998], and can be as much as 1,000 km or more for 10-day trajectories. Clustering algorithms fall into two categories: hierarchical and non-hierarchical. A thorough description of the most widely used clustering algorithms can be found in *Owen* [2003]. The clustering analysis generates groups of similar trajectories, which reveal the most common transport pathways in the dataset. Recently, *Jorba et al.* [2004] used trajectory clustering to generate a flow climatology for Barcelona, Spain, and *Eneroeth et al.* [2003]; *Abdalmogith and Harrison* [2005]; *Sánchez-Ccoyllo et al.* [2006]; *Wang et al.* [2006]; *Borge et al.* [2007] have used clustering analyses to aid in analysis of ground-based measurements for sites in the Arctic, the United Kingdom, Brazil, Europe, and Asia, respectively.

In this work, a climatology of transport pathways is created by performing a clustering analysis for 47 years of atmospheric backward trajectories for the Pico Mountain observatory. This transport climatology is used to analyze the transport to and measurements made at the station during the period of operation from 2001 to 2006. The results are compared to previous findings at the station and are further

explored using the particle dispersion model FLEXPART. Major transport regions and pollutant source regions which affect the observatory and the central North Atlantic are identified, and the seasonal and interannual variability of transport and measurements are described.

Chapter 2

Methods

The trajectory clustering technique used for this research was based on the clustering method originally developed by *Owen* [2003]. The original clustering method and a thorough description of the atmospheric backward trajectories used for clustering are presented in greater detail in *Owen* [2003]. If the reader is unfamiliar with atmospheric trajectories and clustering techniques or seeks a more detailed discussion, we recommend that he or she first read sections 2.1 and 2.2 in *Owen* [2003], as the discussion of clustering and trajectories presented here is less in-depth. To create the final clustering program used here, a few minor changes and one major change were made to the original *Owen* [2003] code; the major change, the application of a non-divergent clustering technique, and its incorporation into the original program are described in this section. All the changes are discussed in greater detail in Appendix A; for more information please read Appendix A after reading this section.

2.1 Trajectories and the Clustering Algorithm

HYSPLIT 4.0 [*Draxler and Rolph*, 2003; *Draxler and Hess*, 1998] was used to calculate kinematic, ten-day backward trajectories using reanalysis meteorological data obtained from *NCEP/NCAR* [2007]. Two groups of trajectories were created. For

use in creating the transport climatology, one group of trajectories was generated at arrival times of 00, 06, 12, and 18 UTC every day from 1960 through 2006. For use in analyzing transport and measurements on an hourly basis using the transport climatology during the period of operation at the Pico Mountain observatory, another group of trajectories was generated every hour from 2001 through 2006. For both groups, sets of six trajectories were calculated around the station at each arrival time. The six trajectories were distributed one degree north, south, east, and west of the station, directly at the station, and 200 m below the station; each set of trajectories is referred to as an ensemble. Each trajectory's location is calculated every hour preceding the station arrival time and at the station.

The non-hierarchical K-means algorithm [Hartigan, 1975] was used for the clustering analysis. The K-means algorithm works iteratively. During each iteration, trajectories are assigned to the nearest cluster center according to distance. In the first iteration, user-defined cluster seeds are used as the initial cluster centers. After all trajectories are assigned, the trajectories in each resulting cluster are averaged to create the new cluster centers for the next iteration. The K-means algorithm was iterated until no trajectories changed cluster assignment in successive iterations. The distance from each trajectory to each cluster center was calculated using the two-dimensional Euclidean distance shown in Equation 2.1. The trajectory latitude and longitude coordinates were converted to north-south and east-west distances in kilometers from the station for this calculation to avoid the exaggeration of distance in the high latitudes that would occur if longitude were used directly.

$$Total\ Euclidean\ Dist. = \sum_{i=0}^{240} \sqrt{((N_{ti} - N_{ci})^2 + (E_{ti} - E_{ci})^2)} \quad (2.1)$$

In Equation 2.1, N is the distance between the latitude coordinate and the latitude at the station, and E is the distance between the longitude coordinate and the longitude at the station, calculated at the latitude coordinate. A 't' indicates the trajectory distance, while a 'c' indicates the cluster center distance, and the distances are calculated

and summed over every upwind hour.

The non-hierarchical method requires the user to specify the number of clusters to be calculated and to provide the initial cluster seeds. Six clusters were selected based on an analysis of the changes in within cluster variance, which is the total distance from each trajectory in a cluster to the cluster center summed over all the cluster centers, and between variance, which is the total distance from each cluster center to the average of all the trajectories summed over all the cluster centers, using different numbers of clusters [Owen, 2003]. A more detailed explanation of within and between variance can be found in Appendix A.1. The cluster seeds can be created randomly or in a separate analysis and were used to initialize the clustering process as discussed above. To create the cluster seeds used here, the trajectories were pre-clustered in a separate analysis, which generated seeds that were more similar to the trajectories than random seeds and reduced the computational time required by the clustering algorithm. For more detail on the cluster seeds used, see Appendix A.3. The final K-means cluster centers were not sensitive to the initial seeds used; the results from the use of random seeds and the pre-clustering step were not significantly different for the 47-yr clustering period. The K-means algorithm for clustering, the use of six cluster centers, the pre-clustering method for cluster seed creation, and the use of Euclidean distance in kilometers were adopted from Owen [2003].

One change was made to the K-means algorithm, which is similar to a unique type of clustering analysis first used by Cape *et al.* [2000], to create the final form used here. In an effort to reduce the effects of trajectory uncertainty, Cape *et al.* [2000] used a new clustering technique in which trajectory ensembles whose trajectories were classified together by the clustering criteria defined core days. The core days were used to help generate the final cluster centers for the study period. We refer to this type of clustering as “non-divergent” clustering because the trajectories in the ensembles do not diverge significantly as they travel backwards in time. More recently, Sánchez-Ccoyllo *et al.* [2006] used Cape’s core day technique with trajec-

tory ensembles for São Paulo, Brazil to analyze the O_3 , CO, and 10 micron or less particulate matter concentrations observed in the city during airflow from different directional quadrants. We have incorporated a version of this non-divergent technique into this clustering analysis for two purposes: one, to reduce the effects of trajectory uncertainty on the transport climatology, and two, to generate cluster centers that are based on times when only one region contributed to the air sampled at the station. As one of the goals of the analysis was to understand the chemical characteristics of the major source regions to the station independently of one another, excluding times of mixed influence is important. Therefore, at the end of each iteration of the K-means algorithm, the cluster assignments of the trajectories in each ensemble were screened; ensembles whose trajectories were assigned to different clusters, indicating the air arriving at the observatory was influenced by more than one region, were classified as mixed-flow. Ensembles whose trajectories were assigned to the same cluster were classified as non-divergent. Only the non-divergent ensembles were used to calculate the cluster centers for the next iteration. By using only non-divergent ensembles, trajectories belonging to mixed-flow ensembles, which may have followed particularly erratic flowpaths and were most likely affected by error, were excluded from the clustering analysis. For more detail on the evaluation of the non-divergent technique, see Appendix A.4. The K-means algorithm with the non-divergent ensemble modification was used to generate four sets of seasonal cluster centers; one each for the 1960 through 2006 springs (March, April, May), summers (June, July, August), falls (September, October, November), and winters (December, January, February). The seasonal divisions follow those used by *Hurrell* [1995] to calculate the seasonal North Atlantic Oscillation index values used in section 3.2.

2.2 Measurement Techniques

CO was determined using a non-dispersive infrared photometer from Thermo Environmental, Inc., model 48C-TL [Val Martín *et al.*, 2006; Honrath *et al.*, 2004]. O₃ was determined by a commercial ultraviolet absorption instrument from Thermo Environmental Instruments, Inc., model 49C. NO, NO₂, and NO_y were determined through O₃ chemiluminescence (NO), conversion via ultraviolet photodissociation (NO₂ to NO), and gold-catalyzed reduction in the presence of CO (NO_y to NO)[Val Martín *et al.*, 2006]. NO_y and NO_x final data were screened as described by Val Martín *et al.* [2007]. Detailed descriptions of the instrument operation, calibration, and precision are found in Honrath *et al.* [2004], Owen *et al.* [2006], and Val Martín *et al.* [2006].

2.3 Upslope Flow Effects at Pico Mountain

The upslope flow impacts at the Pico Mountain observatory were characterized by Kleissl *et al.* [2007]. They found that from October-April the probability of mechanically driven upslope flow causing observation of marine boundary layer (MBL) air was 35–60% per month; this percent represents the upper limit of the frequency of mechanically driven upslope flow as it is based on a conservative calculation of the dividing streamline height upwind of Pico Mountain. Buoyant forcing effects were rare during winter. Frequency of MBL sampling due to mechanically driven upslope flow decreased during May-September to less than 20%. In summer 2004, 40% of the days were affected by buoyant upslope flow, based on measurements made during an intensive mountain-slope sampling campaign that year. Impacts of upslope flow on the O₃ and NO_x observations were extremely small during summer, and no significant diurnal cycles of O₃ and CO were observed. There do not appear to be significant impacts of upslope flow on the summer CO and O₃ measurements at the station. However, the NO_y and NO_x data presented here have been screened for upslope flow

periods, according to the method of *Val Martín et al.* [2007]. The CO and O₃ means presented in section 3.3 were not screened for upslope flow periods. We examined the effects of screening for upslope flow according to the time periods identified by *Kleissl et al.* [2007], and found that the ranking of the cluster centers by mixing ratio would change in only three cases for all the seasons, years, and cluster centers. These changes would not significantly alter the conclusions presented in this thesis. Comparison plots of the screened and unscreened CO and O₃ means can be in Appendix F.

2.4 FLEXPART Simulations

The Lagrangian particle dispersion model FLEXPART [*Stohl et al.*, 2005] generated anthropogenic CO tracer and biomass-burning CO tracer concentrations at the Pico Mountain observatory for use in this analysis. To create the anthropogenic CO tracer, backward FLEXPART simulations (retroplumes) were initialized once every 3 hrs, with particles released on a 1×1 degree \times 500m grid around the observatory over a 1 hr period. The retroplumes were folded with EDGAR version 3.2 emission inventories [*Olivier and Berdowski*, 2001], according to the method of *Seibert and Frank* [2004], to calculate the mixing ratio of CO from several source regions (North America, Europe, Asia, South America, and Africa). The biomass-burning CO tracer concentrations at the station were calculated using FLEXPART forward simulations. For the forward simulations, particles representing fire emissions were released over 3-hour intervals above the locations of active fires, between the surface and 7.5 km with a constant mixing ratio throughout the column [*Lapina et al.*, 2007]. The number of particles released in each column was scaled by the mass of emissions in each grid cell, based on fire emissions estimates from the Boreal Wildland-Fire Emissions Model [*Kasischke et al.*, 2005] for the 2004 and 2005 North American boreal fires. The FLEXPART simulations were driven with European Centre for Medium-Range Forecast (ECMWF)

data, with a resolution of 1×1 degree, 61 vertical levels, available every 3 hours *ECMWF* [2005].

2.5 A Note on Chemistry

The clustering analysis described in section 2.1 is used to analyze transport to and measurements made at the Pico Mountain observatory. Although CO, O₃, NO_y, and NO_x measurements are presented, CO and O₃ are discussed in greater detail. CO is a primary pollutant and is relatively long-lived in the troposphere with a lifetime of 30–90 days; it is destroyed mainly through reaction with the hydroxyl radical [*Seinfeld and Pandis*, 2006]. O₃ is formed through reaction of two groups of primary pollutants, volatile organic carbon (VOCs) and NO_x, and is not directly emitted. The reactions governing the formation and destruction of O₃ are complex, and the tropospheric lifetime varies from days to weeks. However, *Finlayson-Pitts and Pitts* [1986] estimate the tropospheric lifetime to be ~ 10 days. As O₃ is formed by emitted pollutants, it can be transported after being formed or can be formed during transport of an air parcel which contains those precursor emissions. The purpose of the measurement analysis presented in section 3.3 is to generally characterize air which has travelled through the major transport pathways that impact the station; we do not attribute measurements to specific sources, such as the northeastern U.S. or California. In situations where measurements of CO and O₃ are sufficiently high to indicate the presence of relatively recent emissions or production, we attempt to explain these enhancements using known emissions from the most likely upwind source regions on larger scales.

Chapter 3

Results and Discussion

3.1 Transport Pathways to the Pico Mountain Observatory

The non-divergent cluster centers terminating at the Pico Mountain observatory for the four seasons of 1960–2006 are shown in Figure 3.1. The cluster centers show the major flow pathways to the station; each cluster center represents thousands of trajectories whose locations relative to that cluster center vary. To express the variability of the trajectories within each cluster visually, the trajectory endpoints were used to create area-normalized frequency plots on a $1^\circ \times 1^\circ$ grid. Figure 3.2 shows the normalized trajectory endpoint frequency plots for the summer cluster centers, which reveal the extent of the region from which air might be transported by each cluster center. The spring, fall, and winter frequency plots can be found in Appendix B and show transport regions of similar size, relative to the length of the cluster centers. Figure 3.1 also lists the regional labels for the seasonal cluster centers, which are based on the regions identified by the frequency plots.

During summer, the clusters traveled over the following regions: the North Atlantic and Europe (cluster 6), the central and west Atlantic (clusters 1 and 3), the

eastern U.S. (cluster 4), the northern U.S. and southern Canada (cluster 2), and Canada (cluster 5). Transport over these regions was observed during all seasons at the station; however transport over Europe was more likely to occur during spring and summer, given the location of the North Atlantic cluster in those seasons. During fall and winter, the central Atlantic cluster center shifted eastward, and transport from northeastern Africa became more likely. The spring, fall, and winter cluster centers were generally characterized by an increase in long-range transport from the Pacific and across the U.S., relative to summer.

3.2 Seasonal and Interannual Variability of Transport to the Pico Mountain Observatory

We examined the interannual variability of transport during the measurement period at the Pico Mountain observatory by assigning the 2001–2006 hourly trajectory set to the 1960–2006 seasonal cluster centers. The distance method from the clustering algorithm was used to assign the trajectories to the nearest cluster center in a single step. That is, each trajectory was assigned to a cluster center in a single loop. The trajectories were not screened for mixed flow at the end of the loop, so that the percent contribution of transport from each cluster center for each season and year was found. Table 3.1 shows the percent transport from each cluster center by season for 2001–2006.

To identify years with significantly high or low transport, the transport percentages for each year were treated as samples from a larger population for each cluster center. The hypothesis that the years with the minimum and maximum transport percents were outliers from the population was formed and tested using 95 percent distribution intervals for the remaining four years. If the maximum and minimum value fell outside of the distribution intervals, those years had significantly more

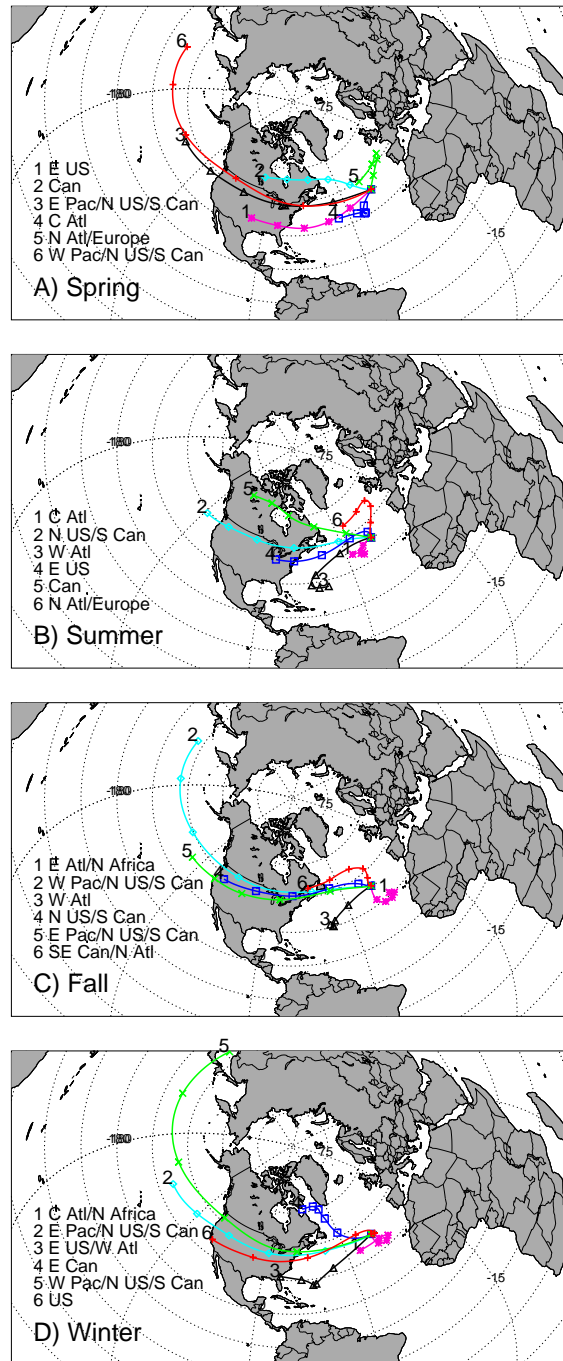


Figure 3.1 1960–2006 Non-divergent Seasonal Cluster Centers. Symbols plotted along the cluster centers show the location every two days preceding arrival at the station. The regional labels for the numbered clusters are listed on the left of each plot. The abbreviated regional names for the cluster centers are defined as follows: E US, eastern U.S.; Can, Canada; E Pac/N US/S Can, eastern Pacific/northern U.S./southern Canada; C Atl, central Atlantic; N Atl/Europe, north Atlantic/Europe; W Pac/N US/S Can, western Pacific/northern U.S./southern Canada; E Atl/N Africa, eastern Atlantic/northern Africa; W Atl, western Atlantic.

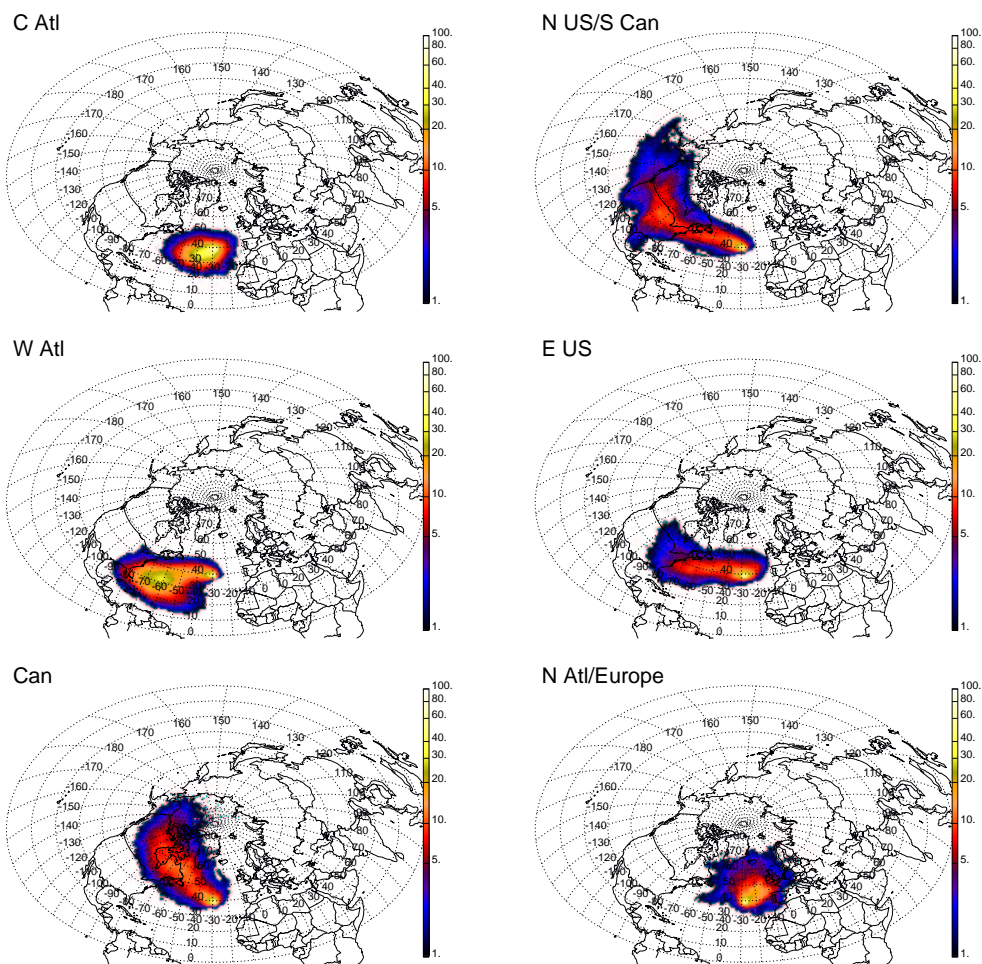


Figure 3.2 1960-2006 Summer Cluster Area-Normalized Frequency. Values less than 1% are not shown. The frequency of each grid cell has been multiplied by the ratio of the grid cell area at the equator to the area of the grid cell of interest, so that the frequencies are not biased by latitude.

or less transport associated with the particular cluster. Using the same hypothesis and outlier test, the maximum and minimum years from the 2001–2006 period were compared to the mean transport from the previous 47-year period, 1960–2006, to determine if the years were unusual relative to transport fluctuations over the longer term. The means and standard deviations used in the outlier tests from the 6-year (2001–2006) dataset and the 47-yr (1960–2006) dataset mean and standard deviation are listed in Table 3.1 as well.

Each season was dominated by flow from different regions, based on the 6 and 47-yr transport percent means. During spring the station generally experienced more transport from the central Atlantic, U.S., and Canada (E US, Can, E Pac/N US/S Can, and C Atl, in Figure 3.1 and Table 3.1) than from the north Atlantic and west Pacific (N Atl/Europe and W Pac/N US/S Can). During summer, transport from the central Atlantic, west Atlantic, and U.S. (C Atl, W Atl, and E US) was greater than transport from Canada (N US/S Can and Can), which was greater than transport from the north Atlantic (N Atl/Europe). Fall was characterized by more transport from the U.S. and Canada (N US/S Can, E Pac/N US/S Can, and Se Can/N Atl) than from the west and east Atlantic (E Atl/N Africa and W Atl), and the least transport from the west Pacific (W Pac/N US/S Can). Winter was similar to fall with the most transport from the U.S. and southern Canada (E Pac/N US/S Can, E US/W Atl, and US), followed by the central Atlantic (C Atl/N Africa), and the least transport from eastern Canada and the west Pacific (E Can and W Pac/N US/S Can). For each season, some of the individual years deviated from the general pattern, as shown by the outlier tests. The most extreme case was during the summer of 2001, which had highly enhanced transport from the western Atlantic (W Atl).

In an attempt to understand the interannual variation, we examined the correlation of the North Atlantic Oscillation index (NAOI) with the yearly transport in the spring and winter cluster from 1960–2006. We selected spring and winter because the amplitude of the NAO is greatest during those seasons [Hurrell, 2001; NOAA/NWS,

Table 3.1 2001–2006 Percent Transport in the 1960–2006 Seasonal Cluster Centers. Outliers from the 2001–2006 distribution at the 95% confidence level are in **bold**, and outliers from the 1960–2006 distribution at the 95% confidence level are underlined. The 6 year mean and standard deviation values calculated without the 2001–2006 maximum and minimum are also shown as the mean \pm the standard deviation. The 47 year mean and standard deviation were calculated including the entire dataset.

	Cluster Center	2001	2002	2003	2004	2005	2006	6 yr	47 yr
S	E US	23	25	19	24	15	20	21 \pm 2.5	23 \pm 4.7
P	Can	24	16	21	15	22	23	21 \pm 3.1	18 \pm 4.8
R	E Pac/N US/S Can	17	27	22	19	11	21	20 \pm 2.1	19 \pm 4.5
I	C Atl	23	22	21	26	27	22	23 \pm 2.0	23 \pm 6.1
N	N Atl/Europe	6.4	3.6	8.3	12	19	12	9.8 \pm 2.9	11 \pm 5.2
G	W Pac/N US/S Can	6.8	6.5	9.1	4.2	4.8	2.3	5.6 \pm 1.3	5.0 \pm 2.6
S	C Atl	17	22	8.8	14	31	25	20 \pm 5.2	22 \pm 6.9
U	N US/S Can	12	8.6	20	16	10	6.4	12 \pm 3.3	13 \pm 4.1
M	W Atl	<u>41</u>	19	31	23	<u>8.3</u>	19	23 \pm 5.6	22 \pm 6.0
M	E US	16	28	19	22	23	19	21 \pm 2.2	21 \pm 5.0
E	Can	11	13	17	17	15	17	16 \pm 1.8	13 \pm 3.7
R	N Atl/Europe	2.5	8.8	4.7	7.0	12	13	8.2 \pm 3.2	8.4 \pm 4.8
	E Atl/Africa	15	9.7	6.4	29	12	17	13 \pm 3.1	17 \pm 7.0
F	W Pac/N US/S Can	4.5	10	4.3	1.5	12	6.9	6.4 \pm 2.7	6.3 \pm 2.9
A	W Atl	11	13	12	13	18	21	14 \pm 2.7	20 \pm 5.4
L	N US/S Can	28	22	28	25	21	23	24 \pm 2.6	22 \pm 3.6
L	E Pac/N US/S Can	20	21	21	<u>8.1</u>	19	19	20 \pm 0.8	18 \pm 5.0
	SE Can/N Atl	22	24	<u>28</u>	24	18	14	22 \pm 3.0	17 \pm 4.9
W	C Atl/Africa	20	12	14	24	30	24	20 \pm 4.9	22 \pm 8.7
I	E Pac/N US/S Can	31	27	33	23	21	34	28 \pm 4.6	28 \pm 6.3
N	E US/W Atl	16	19	18	19	16	9.6	17 \pm 1.6	15 \pm 3.4
T	E Can	9.5	9.7	9.5	14	4.4	4.9	8.4 \pm 2.3	9.9 \pm 5.4
E	W Pac/N US/S Can	17	12	12	4.2	2.9	7.8	8.9 \pm 3.7	9.6 \pm 3.8
R	US	6.8	20	14	17	<u>26</u>	20	18 \pm 3.0	16 \pm 5.0

2005], and the index should be more strongly positive or negative, allowing for more obvious correlations. The NAO indices were taken from *Hurrell* [1995]. Using one-sided linear regression with the NAOI as the independent variable, the NAOI values were compared to the yearly transport percents. During spring and winter, transport from the E Pac/N US/S Can and W Pac/N US/S Can was uncorrelated with the NAOI ($r^2 = 0.01\text{--}0.06$). Transport from Can in spring and E Can in winter was anti-correlated with the NAOI ($r^2 = 0.10, 0.48$, respectively). Spring transport from the C Atl was uncorrelated with the NAOI ($r^2 = 0.04$), but in winter transport from the C Atl/N Africa was correlated ($r^2 = 0.16$). We found spring and winter transport from the E US and E US/W Atl was uncorrelated ($r^2 = 0.00, 0.00$) with the NAOI, but transport from the US in winter was correlated ($r^2 = 0.18$). For the correlated cluster centers in spring and winter, the slopes were found to be significantly different from zero at the 95% confidence level; the slopes in the uncorrelated cluster centers were found not to differ significantly from zero. The NAO appears to be responsible for some of the interannual variation in transport from Can, while transport from the Pacific cluster appears to be independent of the NAO. The relationships for the C Atl and the US with the NAOI varied between spring and winter. The plots of NAOI versus the seasonal transport percents by cluster center can be found in Appendix C.

3.3 Effects of Transport Pathways on the Chemical Observations Made at the Pico Mountain Observatory

To understand the effects of transport pathways identified by the clustering analysis on the chemical observations at the Pico Mountain observatory and the seasonal and interannual variability of those effects, the measurements were analyzed using the clustering results. Hourly averaged O_3 , CO , NO_y , and NO_x data were used.

As performed for the transport analysis, the 2001–2006 hourly trajectories were assigned to the 1960–2006 seasonal cluster centers. However, in this case the trajectory ensembles were screened for mixed flow. During non-divergent arrival times only, when the air sampled at the station was contributed by a single transport region, the measurements were assigned to the clusters. The analysis generated sets of data for each cluster, which were used to identify and compare the effects of different flow pathways on observations at the station and to examine the interannual variability of those observations.

Using only non-divergent trajectories resulted in omission of approximately 60% of the hourly measurement data, but the requirement also provided greater confidence that regions to which the measurements were assigned were the source of the air being sampled. The percent of non-divergent trajectories assigned to each cluster center was not always proportional to the total transport percent of the cluster centers (Table 3.1); in about 50% of the individual years and seasons the ranking of the cluster centers by their total transport percents were different from the ranking by the non-divergent transport percents. As the purpose of the analysis was to investigate the characteristics of transport from potential source regions, having a proportional number of measurements to the total percent transport was not required. However, the reduced number of viable transport and measurement hours had consequences for the clustered data in certain years, which are discussed further below.

Figures 3.3–3.6 show the mean concentrations of species during the seasons of each individual year. During spring (Figure 3.3), transport from the C Atl consistently coincided with the lowest observations of O_3 and NO_y ; the highest O_3 was observed during transport from E Pac/N US/S Can. The NO_x observations during 2004 did not vary significantly between the E US, C Atl, and N Atl/Europe; in 2003, the NO_x observed from the E Pac/N US/S Can was greater than that from the C Atl. The highest CO occurred during transport from Can and the E and W Pac/N US/S Can. The US, C Atl, and the N Atl/Europe were associated with lower CO means.

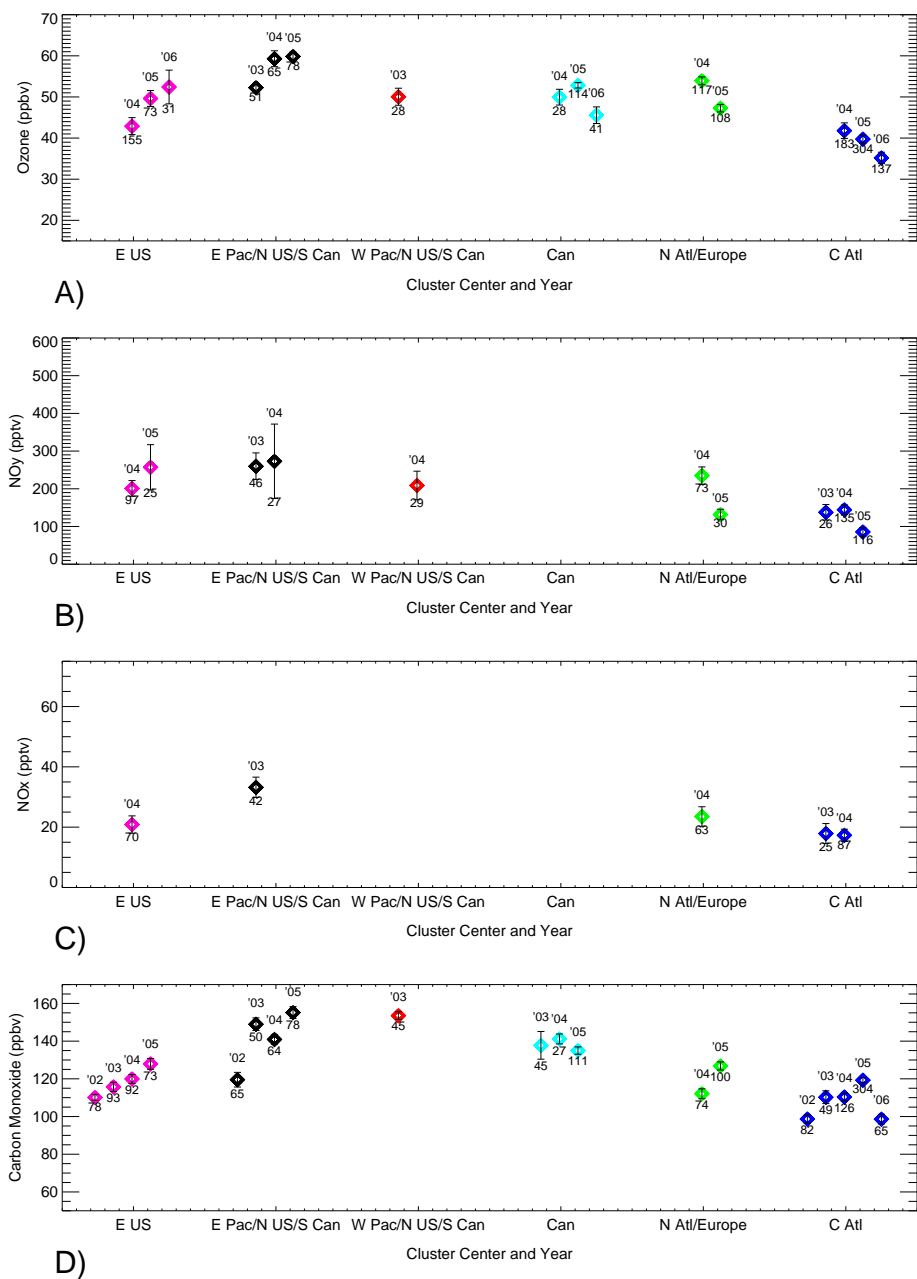


Figure 3.3 Mean Spring Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006. The year is listed above each mean, and the number of points used to calculate each mean is listed below. Means based on fewer than 24 points are not shown. The error bars express the 95 percent confidence interval on the mean.

The difference between the Pac clusters and the US is surprising, as the expected pollutant source for all three clusters is North American anthropogenic emissions. If early boreal wildfire emissions were responsible [Lapina *et al.*, 2006], we would expect to see similar species means between Can and the Pac clusters, which are observed in some, but not all, cases. Two possible explanations for the difference are the faster transport from the North American east coast in the Pac clusters, bringing fresher emissions to the station as compared to the slow speed of the E US cluster, or the influence of export of Asian pollution to the Pacific in spring. Bey *et al.* [2001] found that late winter/early spring is the time of strongest Asian outflow to the Pacific, and Jaeglé *et al.* [2003] found that one-third of the CO sampled over the northeast Pacific was from Asian fuel combustion and biomass burning emissions, based on GEOS-Chem simulations and observations during the PHOBEA-II project. Observations during transport from the Pac clusters could be affected by the presence of Asian outflow in the region, and this possibility is discussed further in section 5 of this chapter.

During summer we generally observed the highest O₃, and CO during transport from the northern clusters; however not all years followed this pattern. In 2001, transport from the W Atl and E US coincided with lower O₃ than the C Atl, and in 2003 O₃ means associated with the W Atl, E US, and Can were not different. For 2004–2006, transport from the E Atl, C Atl, and E US was associated with lower O₃ than the N US/S Can, Can, and the N Atl/Europe. Only the W Atl cluster was associated with low NO_y; the differences between the remaining clusters was less pronounced. For NO_x, however, the C Atl, W Atl, and E US were associated with lower means than the N Atl/Europe, mimicking the observations for 2004–2006 O₃. With the exception of 2005 CO from the W Atl, the yearly patterns of high and low CO followed those observed for O₃. The summer results are discussed further in section 3.4.

During fall, the cluster centers coinciding with the highest and lowest measure-

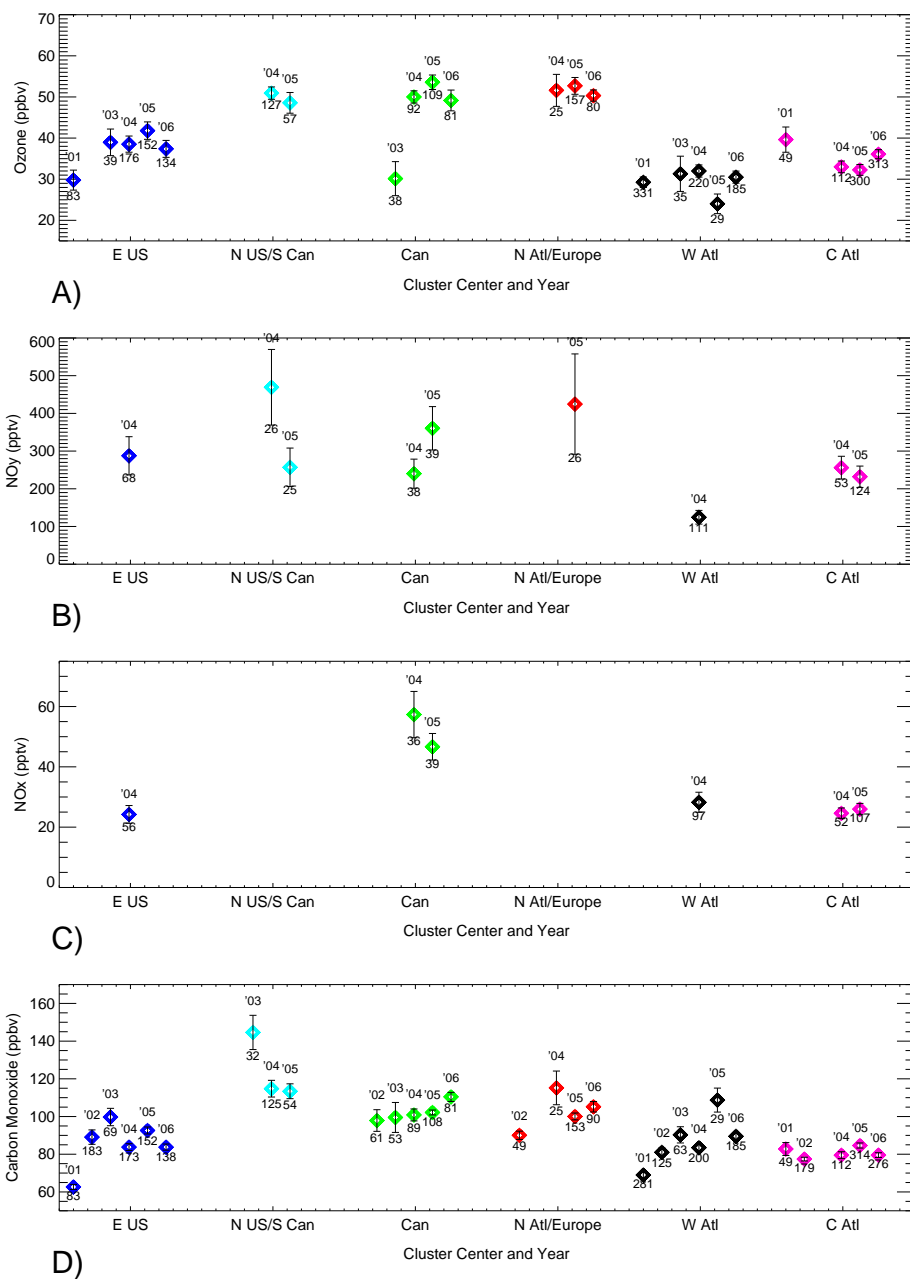


Figure 3.4 Mean Summer Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006. The year is listed above each mean, and the number of points used to calculate each mean is listed below. Means based on fewer than 24 points are not shown. The error bars express the 95 percent confidence interval on the mean.

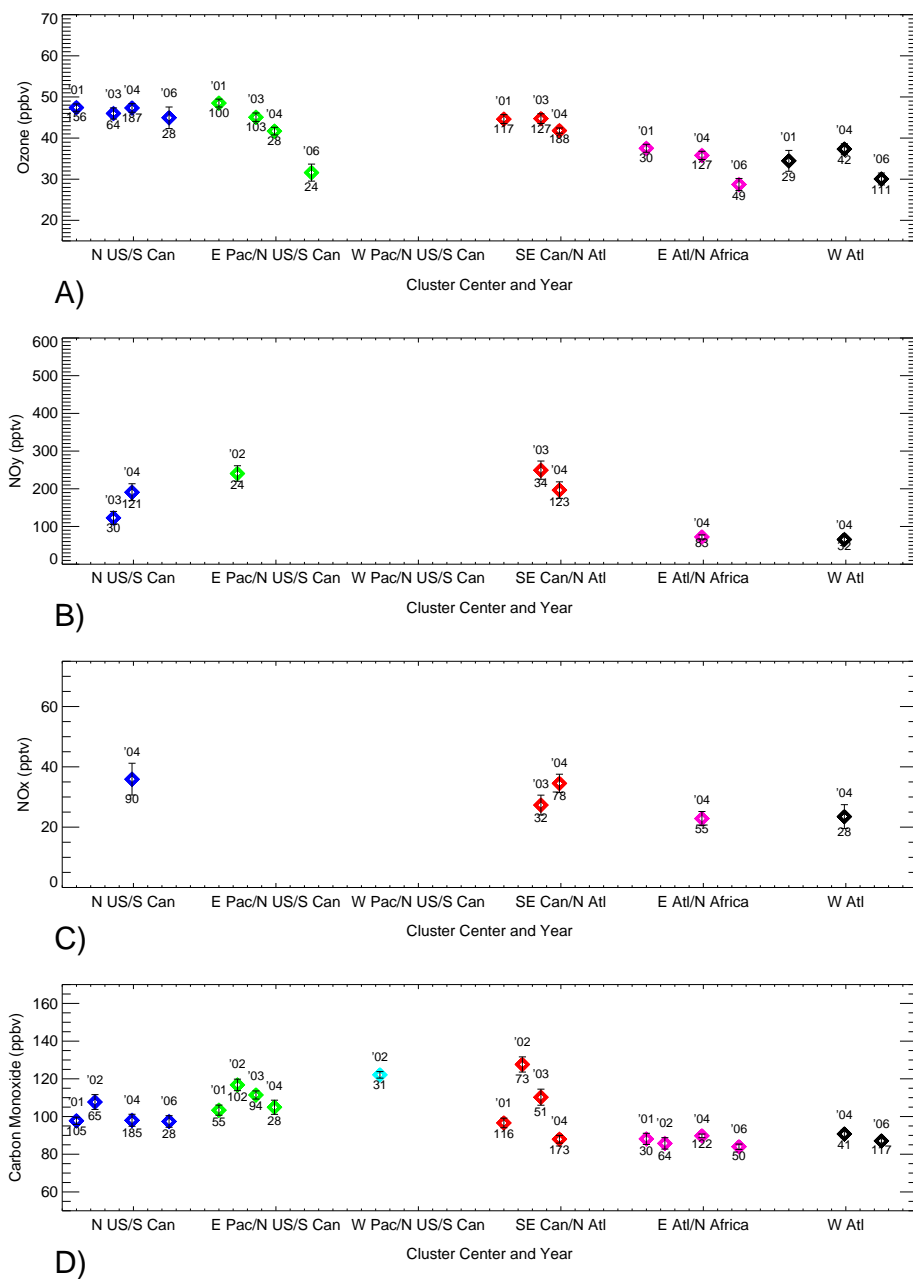


Figure 3.5 Mean Fall Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006. The year is listed above each mean, and the number of points used to calculate each mean is listed below. Means based on fewer than 24 points are not shown. The error bars express the 95 percent confidence interval on the mean.

ments were not consistent between years. For example, 2001 and 2004 had lower O_3 observations during transport from the E Atl/N Africa and W Atl relative to the N US/S Can, E Pac/N US/S Can, and SE Can/N Atl, but this was not the case in 2006, which had higher O_3 from the N US/S Can only. The 2004 NO_y means during transport from the E Atl/Africa and W Atl were lower than the other transport pathways; this difference was less evident in the NO_x means. In 2001, 2002, and 2006, the E Atl/N Africa CO means were lower than the other transport pathways. *Niemi et al.* [2005] performed a ship-based particulate sampling study along the eastern Atlantic from the English Channel to Antarctica during November 1999. While sampling off the coast of northern Africa, they observed high fractions of Saharan dust, as well as biomass burning particulate and sea-salt particulates that may have reacted with anthropogenic or biomass burning emissions. *Li et al.* [2004], using satellite data, found biomass burning emission transport to the Atlantic region off the coast of northwest Africa that may have been sufficient to affect the radiative forcing of the atmosphere there during the low Saharan dust period in November, December (1999), and January (2000). The low species means observed during transport from the E Atl/N Africa indicates that either the pollution whose presence was indicated by these studies was not transported to the observatory from the African coast or the signal was dominated by the clean marine air also transported in that cluster. The slow speed of transport from the E Atl/N Africa may also have played a role in low species means observed.

Poor measurement coverage during winter made drawing general conclusions difficult, but overall, winter was characterized by low variability between clusters (based on the standard deviation of the overall cluster means, relative to spring, summer, and fall). For 2004 and 2005, the C Atl and E US/W Atl were associated with lower O_3 means than the E Pac/N US/S Can and US, but the differences between means were not greater than 10 ppbv. There was adequate NO_y and NO_x data in 2005 to conclude that the NO_y associated with transport from the E US/W Atl and the C

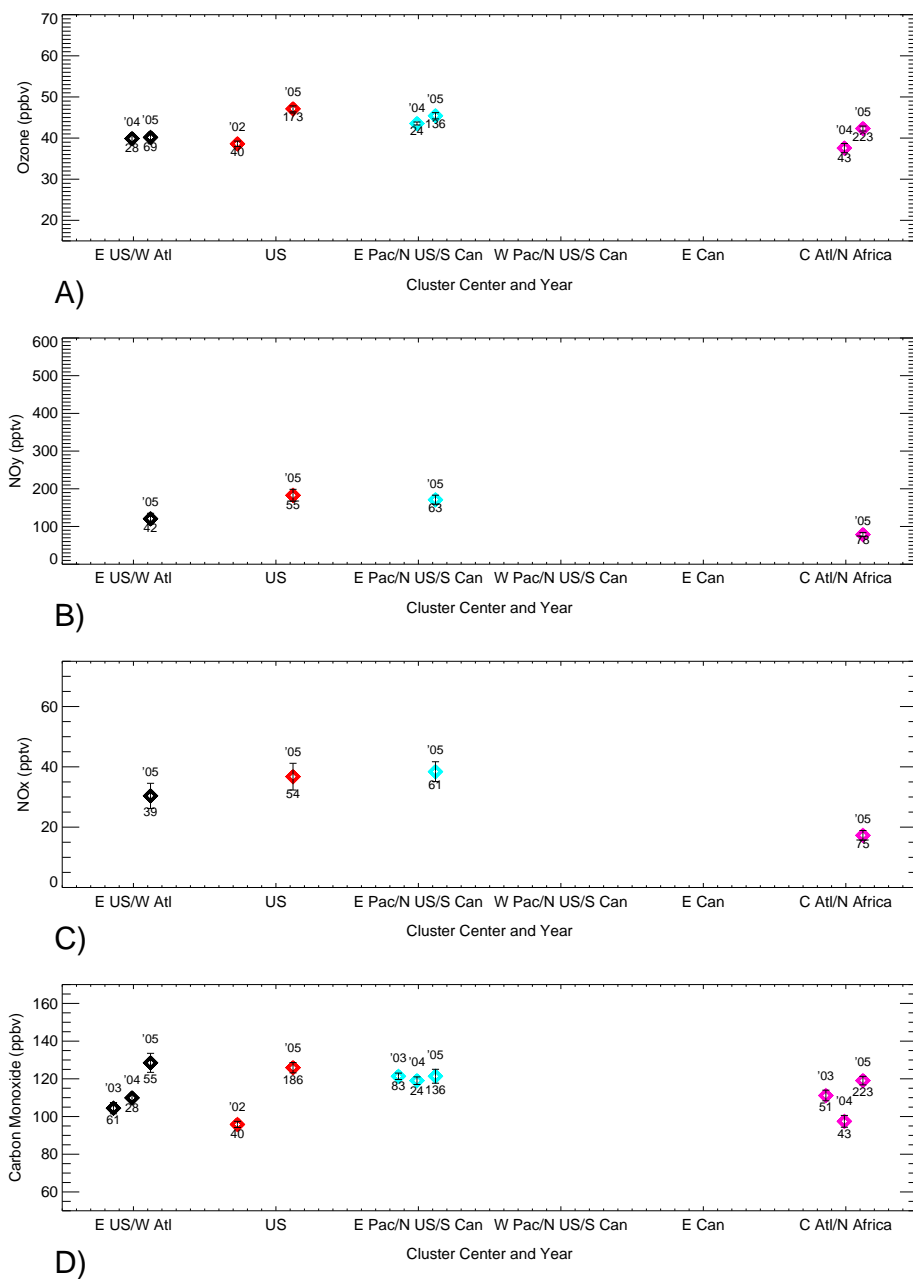


Figure 3.6 Mean Winter Data Mixing Ratios per Cluster Center and Year at the Pico Mountain observatory for 2001-2006. The year is listed above each mean, and the number of points used to calculate each mean is listed below. Means based on fewer than 24 points are not shown. The error bars express the 95 percent confidence interval on the mean.

Atl was lower than that from the E Pac/N US/S Can and the US. Transport from the C Atl coincided with the lowest NO_x measurements. For CO, the years were not consistent.

With few exceptions, the lowest species mixing ratios for all seasons were observed during transport from the Atlantic (from the C Atl in spring, summer, and winter, and the E Atl in fall). The mixing ratios associated with transport from US and W Atl varied, while transport from the Pac clusters tended to coincide with higher observations; in summer the highest means occurred during transport from the northern clusters. Considering all clusters and years, we observed the highest CO and O_3 means during spring, closely followed by summer, and much higher NO_y and NO_x means in summer than in any other season.

3.4 Comparison to Previous Studies at the Pico Mountain Observatory

Several published papers provide analyses for summer observations at the Pico Mountain observatory. The majority of these examine transport events from the U.S. and boreal regions. We compared the previous findings regarding transport and measurements at the station to the results of the measurement analysis with the non-divergent method to determine if these results reflect what was already observed. Where the previous findings and the results presented here did not concur, we have examined those periods in more detail.

Honrath et al. [2004], for 2001–2003, and *Val Martín et al.* [2006], for 2004, concluded that the highest CO observations occurred during biomass burning events and that O_3 was enhanced during these events as well. *Lapina et al.* [2006] examined the entire summer season, rather than individual events, for 2001, 2003, and 2004 and found that CO and O_3 were enhanced during fire periods for each year. 2004 was

a large fire year, especially in Alaska [*Lapina et al.*, 2006], and the 2004 results followed the previous conclusions with enhanced CO and O₃ occurring during transport from all the northern clusters, which have potential to transport air from Alaska and Canada. We observed the same trends during 2005, which was a large fire year in Alaska, an extreme fire year in Siberia and an average fire year in Canada, and during 2006, which was an above-average fire year in Canada and a extreme fire year in Siberia [*NICC*, 2007; *CIFFC*, 2007; *NIFC*, 2007; *Soja et al.*, 2007]. In 2003 however, enhanced CO was observed only during transport from the N US/S Can, and neither the CO nor the O₃ was enhanced during transport from Can. The O₃ associated with transport from Can was significantly lower than that of 2004, 2005, or 2006. Large Siberian fires produced the major biomass burning emissions in 2003, and *Jaffe et al.* [2004] showed transport from the fires was responsible for CO and O₃ enhancements in the northwestern U.S., Alaska, and Canada. *Honrath et al.* [2004] and *Lapina et al.* [2006] showed transport of those emissions in boreal flow from North America reached the station; therefore, we would expect to observe the impacts of those fires from Can in 2003.

Honrath et al. [2004] also concluded that the highest CO levels in 2001 occurred during transport from the U.S. boundary layer. The CO mean associated with transport from the E US observed here was the lowest of any 2001 cluster. This inconsistency and the conflicting observations for 2003 were examined more closely.

To resolve these differences, the *Honrath et al.* [2004] events attributed to the boreal regions of North America in 2003 and to the U.S. in 2001 were compared to the transport hours assigned to Can and the E US, respectively. Only a few hours from the *Honrath et al.* [2004] 2003 boreal biomass burning events were assigned to Can; the remainder were classified as mixed flow. Over half of the transport hours and measurements actually attributed to Can occurred after the end of *Honrath et al.* [2004]’s final event, when O₃ and CO mixing ratios decreased. This indicated that although there were in fact high O₃ observations coinciding with transport from Can,

the method applied here did not recognize them due to the mixed-flow screening. Accordingly, the 2003 O₃ mean associated with Can transport here is similar to the non-fire Canadian mean calculated by *Lapina et al.* [2006] of 35 ppbv. Although transport during fire events was not observed from Can, many of the 2003 transport hours assigned to the N US/S Can coincided with the boreal biomass burning events from *Honrath et al.* [2004], suggesting that the species means associated with transport in that cluster include biomass burning emissions from the region. Upon comparing the 2001 event and transport times, we found that the event hours that *Honrath et al.* [2004] attributed in part or entirely to the U.S. boundary layer were classified as follows: 172 as mixed flow, 32 as W Atl, 36 as C Atl, and 3 as Can, without any assigned to E US, meaning that all of the measurements assigned to the E US occurred during non-event periods. Additionally, 35 of the 49 transport hours assigned to the C Atl occurred during one of *Honrath et al.* [2004]’s high CO events (event 11, identified there as boreal, U.S. boundary layer, and western Atlantic flow), which lasted for several days. This served to enhance the C Atl mean CO, while the E US remained low. These findings show that the non-divergent method used for this analysis limited results in some cases by classifying 60% of the hours as mixed-flow periods, thereby restricting the number of hourly measurements that could be used.

Given this limitation of the non-divergent method, we tested a method of assigning measurements to the clusters that included the mixed-flow periods to determine whether or not the results became more similar to the findings of *Honrath et al.* [2004] and *Lapina et al.* [2006]. At non-divergent times, the coinciding hourly measurements were assigned to the appropriate cluster six times, one for each assigned trajectory; at mixed-flow times, the hourly measurements were assigned to the clusters according to the number of trajectories assigned to each. Therefore, the measurements were assigned to the clusters weighted by the number of trajectories. The treatment was performed for all hours at the station and the means were calculated. Figure 3.7 shows the means for the summer O₃ and CO from both the non-divergent and weighted

mixed-flow methods plotted together.

The inclusion of mixed-flow periods typically caused the clusters and years with elevated non-divergent means to decrease and lower non-divergent means to increase. This is most clear in Figure 3.7 plot A; overall the clusters and years became less different. This effect was not unexpected because, by including mixed-flow times, fire and anthropogenic emissions could be attributed to clean marine flow and vice versa. The 2003 O₃ and CO means from Can, which seemed to be misrepresented by the non-divergent method, were in better agreement with the *Honrath et al.* [2004] and *Lapina et al.* [2006] results when including the mixed-flow periods. The 2001 CO mean for the E US increased slightly, but still did not follow the *Honrath et al.* [2004] conclusion that the highest 2001 CO was observed during transport from the U.S. boundary layer. As U.S. CO enhancements are low relative to those from biomass burning emissions, the high CO events observed previously for the U.S. may have been overwhelmed when averaged with a large number of lower measurements also attributed to transport from the U.S. The improvements of the 2003 O₃ and CO means from Can coincided with the described changes in the majority of the means for the other clusters and years, which were clearly due to mixing. The purpose of the non-divergent requirement was to avoid attributing measurements during pollution events to transport pathways from which they did not originate; if this is not accomplished the means will not represent the air which travels in those pathways. Although exclusion of the mixed-flow periods caused the species means for Can in 2003 to be inaccurate, the cause was easily identified by examining the transport hours in more detail. We therefore concluded that the application of the non-divergent requirement is necessary to ensure that generally the characteristics of the transport pathways are properly represented. More complete measurement coverage would improve the non-divergent method.

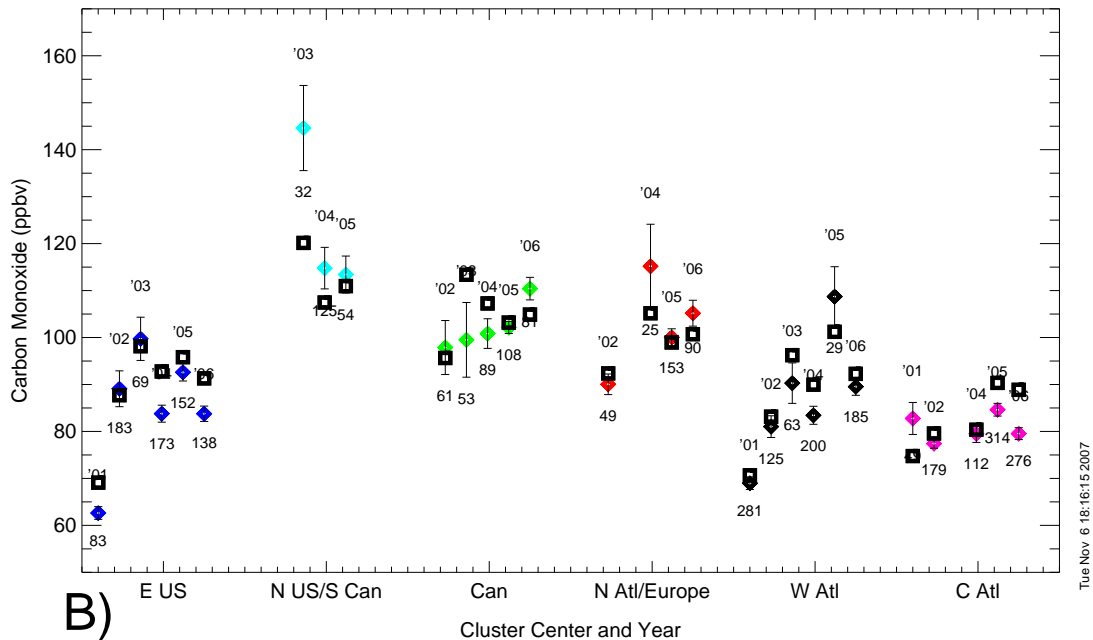
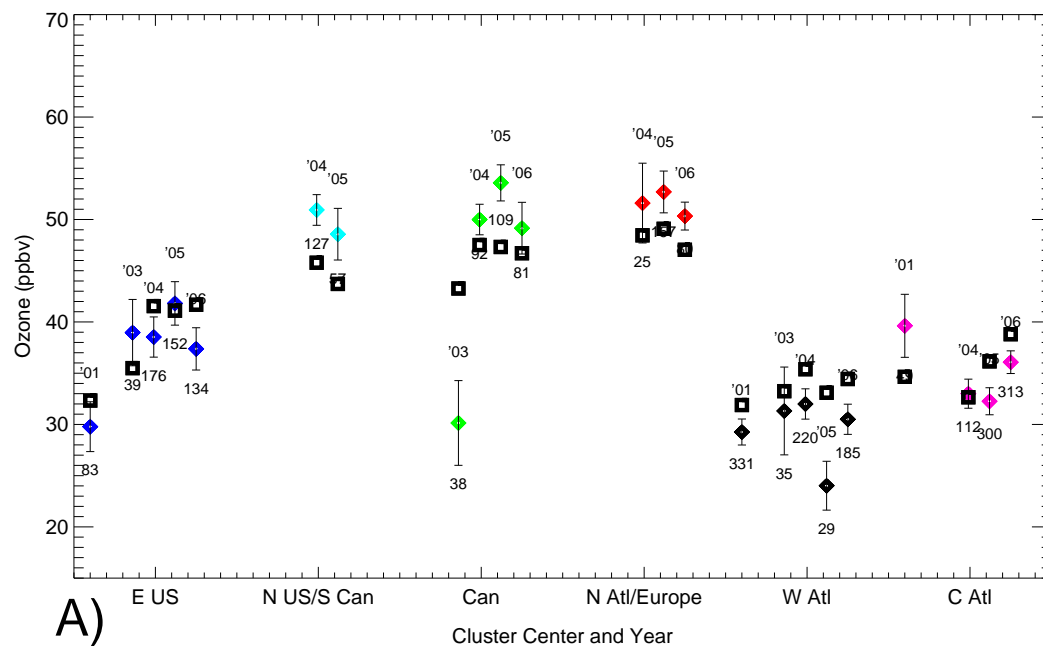


Figure 3.7 2001-2006 Summer Non-divergent and Weighted Mixed-Flow Cluster Mean Comparison. The non-divergent means are the colored diamonds with the 95% confidence intervals, and the weighted mixed flow means are shown in black squares. The year listed above the non-divergent mean corresponds to the mixed flow mean as well, but the number of points listed below the mean corresponds only to the non-divergent mean.

3.5 FLEXPART CO Tracer Analysis for 2003–2005

Spring

In order to investigate the Asian contribution to the observatory in spring and the boreal biomass-burning impacts to the observatory in summer, we used the anthropogenic and biomass-burning (BB) CO tracer values calculated by FLEXPART as described in section 2.4. Using the same procedure as the measurement analysis, the anthropogenic CO tracer concentrations for all the FLEXPART source regions (North America, Europe, Asia, South America, and Africa) were assigned to the seasonal cluster centers for 2002–2004 and the spring and summer clusters in 2005 according to the coinciding hourly cluster assignment of the non-divergent trajectory ensembles. Due to a limited number of measurements during fall and winter of 2005, the FLEXPART CO tracer values were not generated for those seasons. The BB CO tracer, available during the summers of 2004 and 2005, was assigned to the summer clusters in those years by the same method. The means of the assigned tracer values for each year and seasonal cluster center were then calculated. Tables 3.2–3.5 show the mean regional CO contributions to the total anthropogenic CO tracer and the total anthropogenic CO tracer for the listed years and seasons. Table 3.3 also shows the mean BB CO tracer values for the 2004 and 2005 summer cluster centers. The BB CO tracer values are not related to the anthropogenic tracer values and should be considered separately.

According to the anthropogenic CO tracer, the North American region is a significant contributor to the mean total CO in every seasonal cluster center and year. In spring, fall, and winter, the Asian region is the second largest contributor to the mean total anthropogenic CO, most notably in the Pacific clusters, where the Asian contribution can be equal to the North American. The European tracer contributes significantly in the spring and summer N Atl/Europe transport pathway, and in the Atlantic clusters in certain years and seasons. The African contribution, although

generally small relative to the other regions, also increases in the Atlantic clusters in certain years (except for the N Atl/Europe). The South American tracer contributions are never greater than three percent and are frequently zero.

One possible explanation for the enhanced CO and O₃ associated with transport from the Pacific clusters in spring was the increased export of Asian pollution to the Pacific region during spring and subsequent transport to the station. To test this hypothesis, we compare the average Asian tracer contribution in the E Pac/N US/S Can and W Pac/N US/S Can in spring, fall, and winter. If increased Asian pollution was responsible for the enhancements, we would expect to see a larger Asian contribution in spring, relative to fall and winter. We find that the spring E Pac/N US/S Can tracer contribution (24%) lies between the fall (16%) and winter (32%) contributions and that the spring W Pac/N US/S Can Asian tracer contribution (41%) is less than both the fall (67%) and winter (50%) contributions. The Asian fraction of the total anthropogenic CO tracer in the Pacific clusters in spring is not increased relative to fall and winter, and the hypothesis that increased transport of Asian pollution causes the observed enhancements appears to be incorrect.

Earlier we concluded that the enhancements of CO and O₃ in the summer N US/S Can, Can, and N Atl/Europe clusters of 2005 and 2006 were most likely due to increased wildfire emissions from Alaska, Canada, and particularly Siberia; this was based on previous studies at the station, which showed similar boreal wildfire effects in 2003 and 2004. In an effort to support that conclusion we compare the mean BB CO tracer in the 2004 and 2005 summer cluster centers. The highest BB tracer values in 2004 are associated with transport from the N US/S Can, Can and N Atl/Europe, agreeing with previous observations that showed fire impacts during transport from the boreal regions. These tracer means also match the 2004 summer cluster centers with elevated CO means, shown in Figure 3.4. In 2005, transport from the N Atl/Europe is associated with the highest mean BB CO tracer, but the Can mean BB tracer, although greater than that of the C Atl, N US/S Can, and W Atl,

is less than the E US tracer mean. Transport from the N US/S Can is not associated with strong fire emissions. Although 2005 transport from the N Atl/Europe and Can is affected by boreal wildfire emissions, those effects are much weaker than the 2004 boreal wildfire impacts, and are not limited to the northern cluster centers. However, the FLEXPART BB CO tracer alone may not be appropriate for describing fire effects in 2005 because only North American wildfire emissions are included. As stated previously, 2004 was an extreme fire year in Alaska [*Lapina et al.*, 2006], and those emissions were included in the FLEXPART simulations. 2005 was an extreme fire year in Siberia [*Soja et al.*, 2007], but those Siberian emissions, which were more significant in 2005, relative to 2004, were not included in the BB tracer calculation. Therefore, the actual wildfire effects associated with transport in the N US/S Can, Can, and N Atl/Europe in 2005 are most likely larger than expressed by the BB CO tracer alone.

Table 3.2 Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by source for 2003, 2004, and 2005 Springs. The total CO tracer concentration at the station is expressed in the CO column for each cluster and year. The sources are abbreviated as follows: NA=North America, EU=Europe, AS=Asia.

Year	Spring Cluster Centers	Source CO ppb Contribution					Total CO (ppb)
		NA	EU	AS	SA	AF	
2003	E US	8.3	0.8	1.2	0.0	0.4	11
	Can	5.7	1.5	1.4	0.0	0.0	8.7
	E Pac/N US/S Can	8.4	0.4	4.2	0.0	0.0	13
	C Atl	4.0	0.9	0.3	0.0	0.0	5.3
	N Atl/Europe	3.1	2.2	1.5	0.0	0.0	6.9
	W Pac/N US/S Can	14	0.4	8.3	0.0	0.1	23
2004	E US	8.4	0.1	1.5	0.0	0.0	10
	Can	6.7	0.9	1.6	0.0	0.0	9.2
	E Pac/N US/S Can	10	0.2	3.8	0.0	0.0	14
	C Atl	6.6	0.2	1.1	0.0	0.2	8.1
	N Atl/Europe	5.0	8.5	0.7	0.0	0.4	15
	W Pac/N US/S Can	5.3	0.2	5.0	0.0	0.1	11
2005	E US	14	0.1	2.3	0.0	0.0	17
	Can	17	0.9	1.7	0.0	0.0	20
	E Pac/N US/S Can	13	0.2	5.3	0.0	0.1	18
	C Atl	5.3	0.7	0.7	0.0	0.1	6.7
	N Atl/Europe	2.7	4.8	1.3	0.0	0.2	9.0
	W Pac/N US/S Can	7.1	0.5	7.1	0.0	0.1	12
2006	E US	18	0.1	2.0	0.0	0.0	20
	Can	5.5	2.0	1.6	0.0	0.0	9.2
	E Pac/N US/S Can	36	0.1	3.7	0.0	0.0	40
	C Atl	5.4	0.7	0.7	0.0	0.1	7.0
	N Atl/Europe	5.4	4.3	0.7	0.0	0.1	11
	W Pac/N US/S Can	7.9	0.3	3.9	0.0	0.3	12

Table 3.3 Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by source for 2003, 2004, and 2005 Summers. The total CO tracer concentration at the station is expressed in the CO column for each cluster and year. The sources are abbreviated as follows: NA=North America, EU=Europe, AS=Asia, SA=South America, AF=Africa. In addition to the total anthropogenic CO tracer shown in all the seasons, there is an additional summer tracer for 2004 and 2005 which shows the total biomass burning (BB) CO tracer calculated by FLEXPART.

Year	Summer		Source CO ppb Contribution					Total CO	BB Tracer
	Cluster	Centers	NA	EU	AS	SA	AF	(ppb)	CO (ppb)
2003		C Atl	3.8	0.2	0.2	0.0	0.2	4.5	
	2	N US/S Can	11	0.1	1.6	0.1	0.0	13	
	0	W Atl	8.9	0.2	0.5	0.1	0.4	10	
	0	E US	9.1	0.1	0.3	0.0	0.0	9.6	
	2	Can	2.1	0.7	0.8	0.0	0.0	3.6	
		N Atl/Europe	4.4	0.3	0.1	0.0	0.1	5.0	
2004		C Atl	4.0	0.3	0.3	0.0	0.1	4.7	
	2	N US/S Can	18	0.1	1.5	0.0	0.0	19	
	0	W Atl	8.3	0.0	0.5	0.1	0.2	9.0	
	0	E US	7.1	0.0	0.5	0.0	0.3	7.9	
	3	Can	6.6	0.2	1.1	0.0	0.1	8.0	
		N Atl/Europe	3.6	3.5	0.1	0.0	0.0	7.2	
2005		C Atl	5.4	0.1	0.0	0.0	0.0	5.6	6.9
	2	N US/S Can	14	0.1	1.5	0.0	0.0	15	27
	0	W Atl	8.8	0.1	0.3	0.2	0.2	9.6	2.6
	0	E US	10	0.1	0.4	0.0	0.0	10	6.9
	4	Can	8.3	0.4	0.9	0.0	0.0	9.6	8.7
		N Atl/Europe	2.3	4.5	0.2	0.0	0.0	7.1	11
2006		C Atl	5.0	0.4	0.1	0.0	0.2	5.7	3.0
	2	N US/S Can	24	0.1	0.8	0.0	0.0	25	0.6
	0	W Atl	7.6	0.1	0.7	0.1	0.1	8.7	0.4
	0	E US	12	0.5	0.4	0.1	0.0	13	5.5
	5	Can	4.9	0.2	0.8	0.0	0.0	5.9	4.1
		N Atl/Europe	4.9	4.8	0.1	0.0	0.3	10	6.1

Table 3.4 Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by source for 2002, 2003, and 2004 Falls. The total CO tracer concentration at the station is expressed in the CO column for each cluster and year. The sources are abbreviated as follows: NA=North America, EU=Europe, AS=Asia, SA=South America, AF=Africa.

Year	Fall Cluster Centers	Source CO ppb Contribution					Total CO (ppb)
		NA	EU	AS	SA	AF	
2002	E Atl/N Africa	4.5	0.2	0.8	0.0	0.8	6.3
	W Pac/N US/S Can	4.7	0.5	5.2	0.0	0.1	11
	W Atl	6.6	0.1	1.0	0.2	0.1	8.1
	N US/S Can	6.1	0.3	1.8	0.0	0.1	8.2
	E Pac/N US/S Can	19	0.1	2.0	0.0	0.0	21
	SE Can/N Atl	4.6	4.0	0.6	0.0	0.0	9.3
2003	E Atl/N Africa	3.8	0.7	0.5	0.0	0.3	5.2
	W Pac/N US/S Can	2.4	0.1	4.3	0.0	0.1	6.9
	W Atl	4.9	0.1	1.0	0.0	0.1	6.2
	N US/S Can	8.8	0.1	1.6	0.1	0.0	11
	E Pac/N US/S Can	10	0.4	2.8	0.0	0.0	14
	SE Can/N Atl	5.9	1.0	0.7	0.0	0.1	7.7
2004	E Atl/N Africa	2.3	2.2	0.3	0.0	2.3	7.1
	W Pac/N US/S Can	0.6	0.1	5.1	0.0	0.0	5.8
	W Atl	4.0	0.7	0.7	0.0	0.3	5.6
	N US/S Can	6.3	0.2	2.2	0.0	0.0	8.8
	E Pac/N US/S Can	11	0.5	2.4	0.0	0.0	14
	SE Can/N Atl	6.1	1.7	0.4	0.0	0.0	8.2

Table 3.5 Contribution to Total FLEXPART CO Tracer at the Pico Mountain observatory by source for 2002, 2003, and 2004 Winters. The total CO tracer concentration at the station is expressed in the CO column for each cluster and year. The sources are abbreviated as follows: NA=North America, EU=Europe, AS=Asia, SA=South America, AF=Africa.

Year	Winter Cluster Centers	Source CO ppb Contribution					Total CO (ppb)
		NA	EU	AS	SA	AF	
2002	C Atl/N Africa	6.2	5.0	0.9	0.0	1.3	13
	E Pac/N US/S Can	6.3	0.2	4.4	0.0	0.1	11
	E US/W Atl	8.2	0.3	1.5	0.1	0.1	10
	E Can	4.4	1.7	1.8	0.0	0.0	7.9
	W Pac/N US/S Can	6.3	0.5	6.2	0.0	0.1	13
	US	7.7	0.2	1.7	0.1	0.0	9.7
2003	C Atl/N Africa	4.7	1.0	1.6	0.0	0.0	7.3
	E Pac/N US/S Can	6.8	0.3	3.5	0.0	0.1	11
	E US/W Atl	6.0	0.1	1.4	0.0	0.0	7.6
	E Can	9.4	2.8	3.5	0.0	0.0	16
	W Pac/N US/S Can	3.5	0.5	6.4	0.0	0.1	10
	US	3.5	0.1	2.8	0.0	0.0	6.4
2004	C Atl/N Africa	3.1	5.9	0.4	0.0	1.2	11
	E Pac/N US/S Can	12	0.2	3.7	0.0	0.0	16
	E US/W Atl	15	0.1	1.3	0.0	0.0	16
	E Can	8.5	0.8	1.2	0.0	0.0	11
	W Pac/N US/S Can	9.6	0.2	6.5	0.0	0.1	16
	US	4.7	1.4	1.5	0.0	0.0	7.6

Chapter 4

Summary and Conclusions

The non-divergent, K-means clustering technique applied to the 40-year trajectory dataset for the Pico Mountain observatory was used to create sets of 6 cluster centers for each season. The seasonal cluster centers identified the U.S., Canada, the Atlantic, the Pacific, Europe and eastern Africa as the major regions over which air was transported before arriving at the station. By combining the seasonal cluster centers with all 2001–2006 hourly trajectories, the percent of time the station was influenced by each region for each year and season was found. Transport over the U.S., Canada, and the Atlantic most frequently affects the observatory. The interannual variation of spring and winter transport from Canada appeared to be controlled in part by the NAO. Spring and winter transport from the Pacific was poorly correlated with the NAO.

The cluster centers were also used to analyze the hourly observations of O_3 , CO , NO_y , and NO_x made at the station. Measurements were assigned to the clusters based on the assignment of the coinciding hourly trajectory ensemble; only non-divergent trajectory assignments were used, which reduced the influence of mixed flow and enabled us to study the effects of the individual transport regions at the station. Species means were calculated for each year, season, and cluster center. In spring,

we observed the lowest O_3 and NO_y from the C Atl, and the lowest CO from the C Atl, the N Atl/Europe, and the E US. The high CO measurements associated with transport from the Pac clusters, the highest observed during any season, were thought to possibly be affected by strong Asian outflow to the Pacific in spring; this conclusion was disproven by the FLEXPART simulations, which showed no increase in Asian contribution to the total anthropogenic CO tracer during transport from the Pac clusters, relative to fall and winter. During summer of 2004, the highest CO and O_3 observed occurred during transport from the N US/S Can, Can, and the N Atl/Europe, agreeing with previous studies of boreal wildfire events observed at the station. We continued to observe the impacts of boreal fires during 2005 and 2006. Although, the FLEXPART BB CO tracer showed weaker fire impacts in 2005 than 2004, the tracer may not fully describe 2005, as Siberian fire emissions were not included. The NO_y and NO_x means observed during summer were significantly higher than observations during any other season.

Comparison of the results of this paper with those of other analyses for the observatory revealed some limitations of the method. The mixed-flow screening in combination with the data screening severely minimized the number of measurements used in the analysis. In the case of 2001 and 2003, this led to misrepresentation of the chemical characteristics of the U.S. and Canada, respectively, as compared to previous observations. However, inclusion of mixed-flow periods when calculating the cluster means led to obvious effects of mixing on the final mean values for the remaining cluster centers, and avoiding those mixing effects was the purpose of the non-divergent method. Improved measurement coverage would help make the non-divergent method accurate for all clusters, years, and seasons.

Through the use of the non-divergent technique for clustering, transport, and measurement analysis we identified the major transport pathways impacting the Pico Mountain observatory and quantified their transport frequency and chemical characteristics. As measurements will continue to be made at the station, the method may

be useful for future analyses there, and perhaps as an alternative for atmospheric transport studies at other air monitoring sites as well.

Chapter 5

Future Work

Additional development of some of the results and conclusions will be necessary. Possible changes include additional analysis of the effects of the NAO and analysis of the effects of interannual variability of transport on the measurements. The effects of the NAO on transport frequencies for the measurement period at the station and the longer-term clustering period were already examined. Comparison of these results with the studies of the NAO impacts at other ground-based air monitoring stations in the region to determine whether our findings coincide would be useful and interesting.

Investigation of the dependence of interannual measurement variability on the interannual variability of transport from the seasonal cluster centers would link the two major sections of results from this research (transport and measurements). Some preliminary analysis of the correlations between CO means and the transport percents in spring and summer has been performed; however, the results were inconsistent. In spring we found transport and measurements to be correlated from the Atlantic clusters and anti-correlated from the U.S. clusters, but in summer the correlations were opposite and weaker. To draw clear conclusions about the effects of transport variability on the measurements, additional analysis is required; however, the limited number of data points available may make analysis difficult.

Analyses of the type performed here could be applied to the 2008 and 2009 summer measurement campaigns planned for the observatory. If the measurements from those summers are analyzed using the cluster centers calculated here, the additional means might enable an investigation of regional pollution trends, which were not examined in this thesis. There has been active research on pollutant trends in the eastern U.S., and comparison of those observed trends with any trends observed at the station may be useful and interesting.

Another application of these results would be a study of the seasonal cycles of O_3 and CO and the impact of those cycles on the observed species concentrations from the seasonal transport pathways. Seasonal differences in species means were only briefly mentioned here, but spring and summer maxima of O_3 measurements have been identified at many North Atlantic and western European measurement sites. *Val Martín et al.* [2007] analyzed the seasonal cycles of NO_y and NO_x at the station, and quantification of the O_3 and CO cycles at the station would compliment the existing body of work well.

One important feature of atmospheric transport was not addressed in this research: chemistry within the transported pollution plumes and major transport pathways. The chemical transformations of species that occur during transport to the station have been explored in previous studies at Pico Mountain and will continue to be investigated in the future. Formation and destruction of O_3 and the lifetime and transformation of NO_y and NO_x are of particular interest. Although the emissions from a particular region play a large role in what is measured at the station, the complexities and variations of chemistry during transport also control downwind concentrations. To better understand the observations at the station during transport from the major source regions, the chemistry occurring in the different transport pathways needs to be characterized.

Appendix A

Additional Detail on Analyses to Determine the Final Clustering Method

Section 2.1 describes the final clustering method used for this research. To arrive at the final version, several modifications were made to the original clustering technique of *Owen* [2003]. This appendix gives a brief background on cluster variance and describes the major program changes and analyses of cluster seeds and the non-divergent method. The reasons for selecting the non-divergent clustering method are summarized.

A.1 Describing Clusters: Variance

Variance was used to describe the cluster center sets and to compare the results of the various analyses. Three different types of variance were calculated for each cluster center set: between cluster variance, within cluster variance, and total variance.

Between cluster variance describes how different the cluster centers are from the average trajectory. The average trajectory is the trajectory created by averaging the

endpoints of all of the trajectories. Equation A.1 shows the calculation of between cluster variance. The distance between each point on the average trajectory and a cluster center is calculated (d_{ik}) and squared. The distances are summed over the number of endpoints and then over the cluster centers.

$$\textit{Between Cluster Variance} = \sum_{k=0}^5 \sum_{i=0}^{240} (d_{ik})^2 \quad (\text{A.1})$$

Larger values of between cluster variance indicate that the cluster centers are further away from the average trajectory. This also indicates that the cluster centers are more different from each other. Dissimilar cluster centers represent transport pathways that are also farther apart; therefore, large between cluster variance is desired.

Within cluster variance describes how well each cluster center represents its member trajectories. The total squared distance between each point on a member trajectory and its cluster center (d_{imk}) is summed over all the endpoints and all the member trajectories. The values for each cluster center are then totaled. Equation A.2 shows the calculation; N_{MT} is the number of trajectories assigned to each cluster center.

$$\textit{Within Cluster Variance} = \sum_{k=0}^5 \sum_{m=0}^{N_{MT}} \sum_{i=0}^{240} (d_{imk})^2 \quad (\text{A.2})$$

Smaller values of within cluster variance indicate that the member trajectories are closer and more similar to their assigned cluster centers. Therefore, the cluster centers are more representative of their trajectories. In section 4 of this appendix we compare the non-divergent and mixed-flow clustering methods, and the number of trajectories used by each method varied. If a method uses fewer trajectories, a smaller value of within cluster variance is inherently expected. To allow for comparison between methods in such cases, normalized within cluster variance was used. Normalized within cluster variance is within cluster variance divided by the number of trajectories used.

Total variance is the sum of the distance between each endpoint on the average trajectory and each endpoint on every trajectory (d_{im}) squared. It does not describe

the quality of the cluster centers and was not used to compare methods, but is shown in analysis tables as it was used to verify that the programs were running properly. If two methods used the same trajectory set, then total variance should be equivalent between the methods.

$$Total\ Variance = \sum_{m=0}^{N_T} \sum_{i=0}^{240} (d_{im})^2 \quad (A.3)$$

Between and within cluster variance were used to compare the cluster center sets; total variance was used for code verification as it only describes the trajectories. When comparing two methods using variance, the difference between the methods is expressed as a percent.

A.2 Major Changes to the Owen 2003 Clustering Code

In addition to the non-divergent modification described in section 2.1, three other major changes were made to the *Owen* [2003] clustering code. These changes include modification of the program that converts trajectory coordinates into distances, modification of the program that loads the trajectory data for clustering, and modification of the manner in which the large data arrays are handled within the clustering program. In the program, a routine is called to convert trajectory coordinates to distance in kilometers. The original version left out a keyword necessary for the distances to be calculated in the north-south and east-west directions at the proper latitude. In the final version, the program was modified and this keyword was included. In addition to providing the coordinates of the trajectory locations, the HYSPLIT model provided many other variables containing information about the trajectory endpoints. These extra variables were not used in the clustering program, therefore loading them wasted computational time and resources. The trajectory loading program was modified so that these variables are now excluded before being loaded, rather than after.

This change and the changes to the array manipulations in the clustering program save computing time. The change to the trajectory coordinate conversion routine made the final cluster centers more accurate.

A.3 Seed Sensitivity Analysis: Methods and Results

Owen [2003] created a two-step, non-hierarchical, K-means clustering procedure to generate the initial cluster seeds that are sent to the final K-means clustering process. The first step of the seed procedure used randomly generated cluster seeds to begin clustering and updates the cluster centers after every trajectory assignment. The cluster centers from the first step are sent to the second step, which is similar to the final K-means process, but clusters the trajectories using a simplified distance equation. The cluster centers from the second step are used as the seeds for the final clustering process. The purpose of this seed-generating procedure was to reduce the run time for the final K-means process, which was longer when random initial seeds were used.

While modifying and testing the original code, as was described above, often only a few years or less of the trajectories were used during test runs to save time. The final cluster centers resulting from these test runs were found to vary, depending on the type of cluster seeds used to initialize the final clustering step. This sensitivity of the results prompted an analysis of the different cluster seeds and the resulting cluster centers over the long-term clustering period.

For this analysis, three permutations of the seed-generating procedure were compared. In the first variation, the second step of the process was stopped when 2,000 trajectories or fewer changed cluster centers during an iteration. The cluster centers at the end of that iteration were sent to the final clustering process as seeds. In

Table A.1 1960-2006 Seasonal Seed Sensitivity Analysis of Cluster Variance.

Season	Seed Type	Norm. Within Variance	Between Variance	Total Variance	% Diff. NWV	% Diff. BV
Spring	Convergent	23094	277689	1.64×10^9	0	0
	2,000 Traj	23669	251968	1.77×10^9	2.49	-9.26
	Random	23094	277566	1.64×10^9	< 0.001	< 0.001
Summer	Convergent	17803	176832	1.36×10^9	0	0
	2,000 Traj	17803	176832	1.36×10^9	0	0
	Random	17800	176777	1.36×10^9	< 0.001	< 0.001
Fall	Convergent	22004	271813	1.61×10^9	0	0
	2,000 Traj	22004	271813	1.61×10^9	0	0
	Random	21988	272060	1.60×10^9	< 0.001	< 0.001
Winter	Convergent	26428	306453	2.15×10^9	0	0
	2,000 Traj	26428	306453	2.15×10^9	0	0
	Random	26428	306453	2.15×10^9	0	0

the second variation, the second step of the process was stopped when 0 trajectories changed cluster centers (convergence). In the third variation, the final clustering step was initialized using random seeds. The within cluster, between cluster, and total variance of the final cluster centers resulting from each variation are shown in Table A.1 for the 1960–2006 seasons. The percent difference was calculated relative to the convergent seed variance. For the normalized within variance, this percent difference is abbreviated as % Difference NWV, and for between variance it is abbreviated as % Difference BV. In Table A.1 the total variance for each seed type differed in certain seasons because the non-divergent method clustering was used and resulted in slightly different trajectory sets being used to calculate the final cluster centers. Figures A.1 through A.6 show the final cluster centers for the spring and summer seasons.

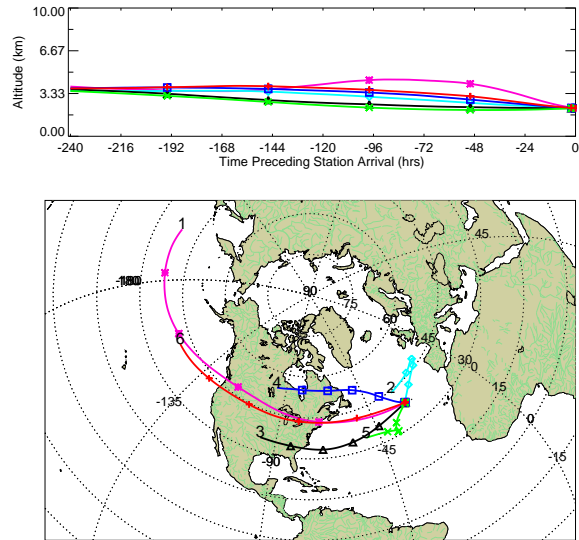


Figure A.1 1960-2006 Spring Cluster Centers Produced with Convergent Seeds.

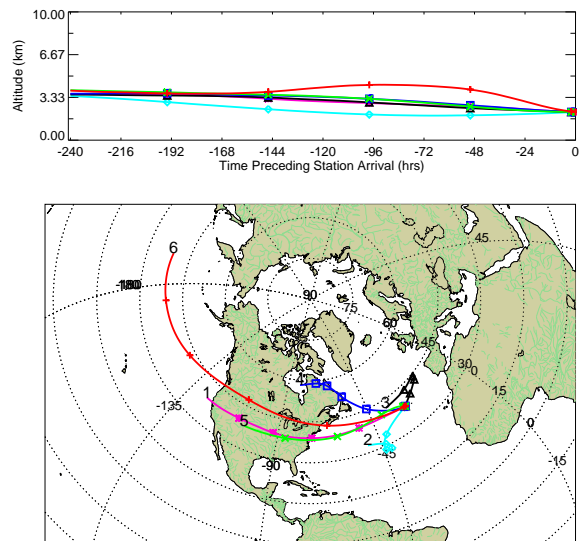


Figure A.2 1960-2006 Spring Cluster Centers Produced with 2,000 Trajectory Limited Seeds.

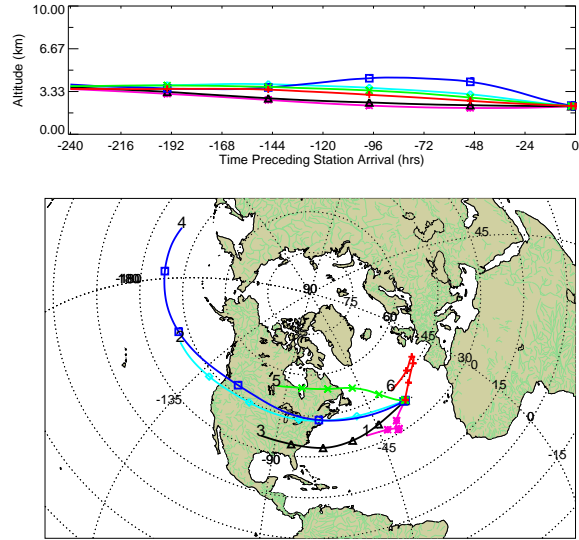


Figure A.3 1960-2006 Spring Cluster Centers Produced with Random Seeds.

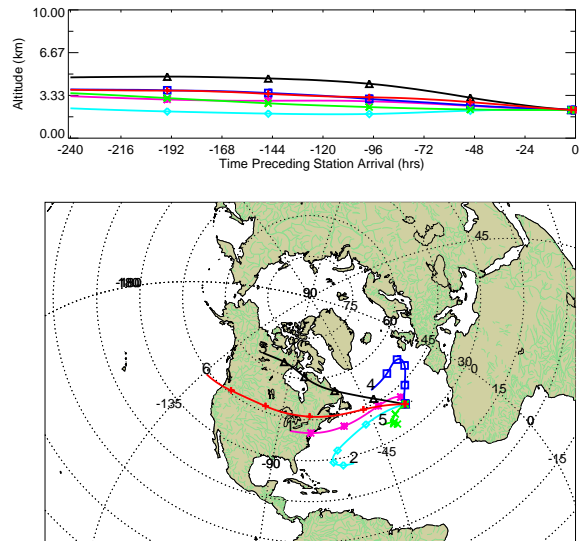


Figure A.4 1960-2006 Summer Cluster Centers Produced with Convergent Seeds.

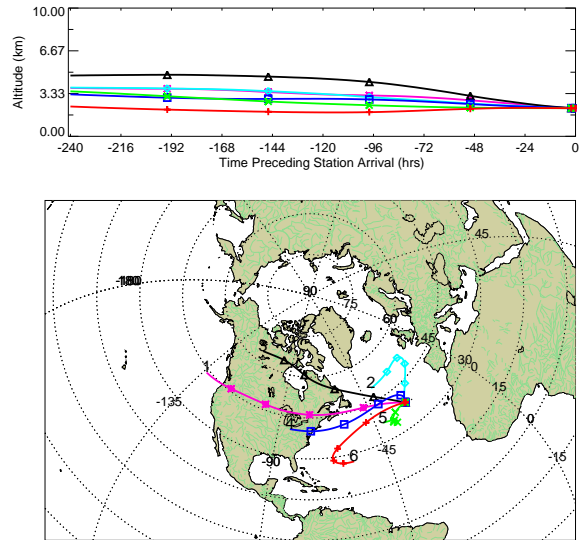


Figure A.5 1960-2006 Summer Cluster Centers Produced with 2,000 Trajectory Limited Seeds.

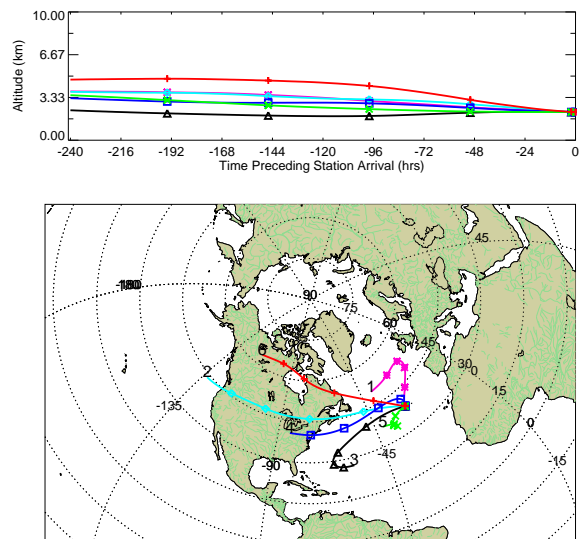


Figure A.6 1960-2006 Summer Cluster Centers Produced with Random Seeds.

As seen in Table A.1, the only difference in the variance between the final cluster sets was for the spring within and between cluster variance. All of the other seasons were nearly identical for all seed types. The differences between the seed types evident in the variance was also observed in the final cluster plots for spring. The random and convergent seeds produced final cluster centers in very similar locations (the colors and numbers of the cluster centers are not important for this comparison), but the locations of the cluster centers for the 2,000 trajectory limitation seeds were noticeably shifted, particularly over Canada and the continental U.S. In spite of the difference in locations, the regions over which the 2,000 trajectory limitation cluster centers traveled were not significantly different than those regions for the random and convergent cluster centers. The summer cluster plots show that the final cluster centers for all seed types were quite similar, which was also the case in fall and winter. As the three seed types for summer, fall, and winter produced the same results, and the differences in spring were not severe enough to change the transport regions, we concluded that the clustering program was not significantly sensitive to the seeds over the long term period used in this analysis. Therefore, for the final clustering program, the convergent seed method was used, following *Owen* [2003]’s method.

A.4 Clustering Method Final Decision: Mixed Flow versus Non-divergent

As discussed in section 2.1, the final clustering technique used for the analysis was the non-divergent clustering method. To arrive at this decision, both non-divergent and mixed-flow methods were compared, so that the technique that created the most appropriate cluster centers for the research goals was used. The mixed-flow clustering method is the *Owen* [2003] clustering code with the changes described in section 2 of this appendix without the non-divergent modification. The main benefits of the

non-divergent technique are the generation of cluster centers that represent air flow pathways during times when the air at the station came from a single region and a reduction of the effects of trajectory uncertainty on the final cluster centers. The non-divergent requirement reduces the effects of trajectory uncertainty by using only non-divergent ensembles; in non-divergent ensembles, the trajectories typically follow very similar pathways, and the possibility that all six trajectories have the same amount of error is low. The major disadvantage of the technique is that 60% of the trajectories are not used for clustering. Mixed-flow clustering uses all of the trajectories, but the final cluster centers represent airflow pathways from multiple potential source regions. The cluster centers generated by both methods are compared in the first portion of this analysis. The results are then combined with a logical discussion of how well the methods suite the research goals.

The 1960–2006 seasonal mixed flow and non-divergent cluster centers are shown in Figures A.7 and A.8. The cluster centers for the two methods differed the most during spring, and somewhat less during the other seasons. Although the clusters varied, generally transport from the same regions was indicated by both methods for each season; the only exception was the north Atlantic cluster in spring and summer. The non-divergent north Atlantic clusters were located farther northeast than the mixed flow clusters, making European transport a possibility. These differences are reflected in the variance values, which are shown in Table A.2. The percent difference for normalized within variance (% Diff. NWV) and between variance (% Diff. BV) were calculated by subtracting the non-divergent value from the mixed flow value and dividing by the non-divergent value.

The variance comparison indicated that the non-divergent method created the more representative and unique cluster center set. In all the seasons, the non-divergent clusters have a smaller normalized within variance and larger between variance; with the exception of normalized within variance for summer, the differences are always greater than 5%, particularly for between variance. In addition to statistical supe-

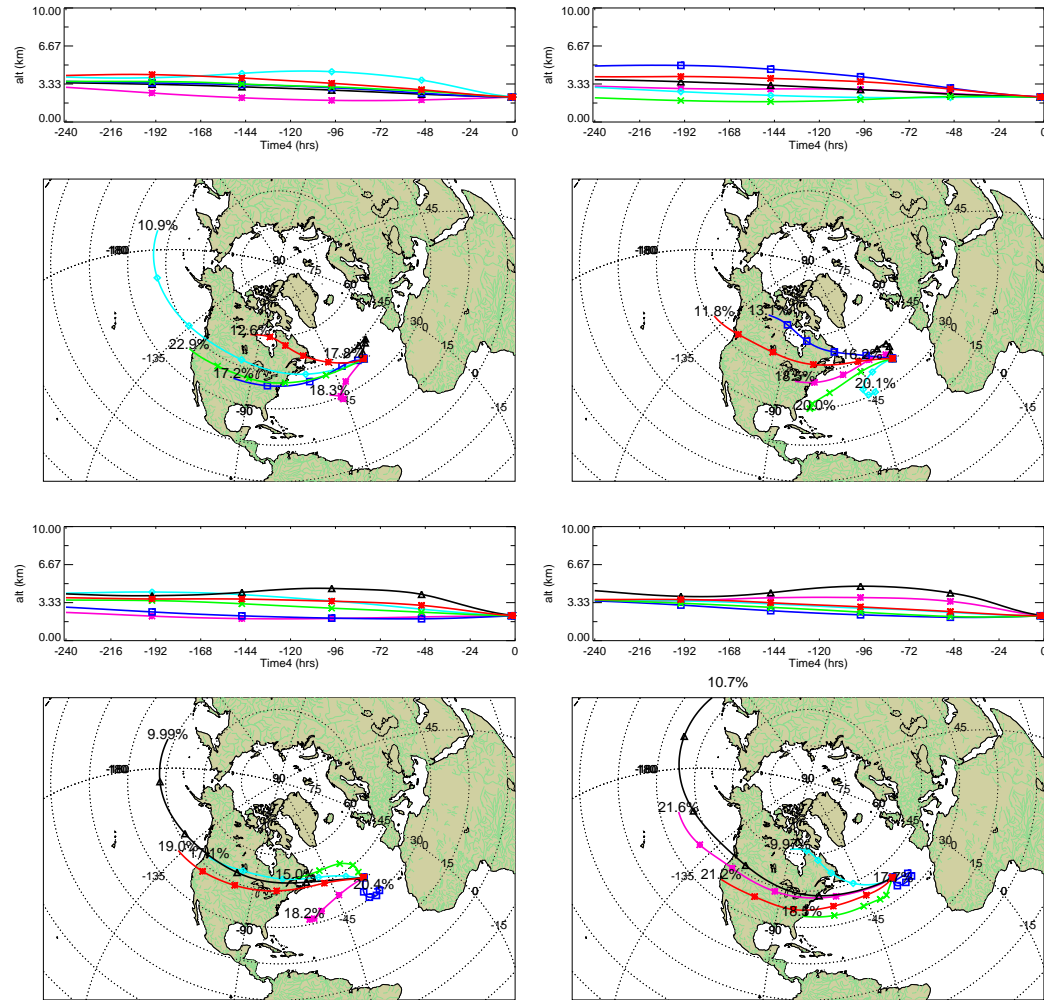


Figure A.7 1960–2006 Mixed-Flow Seasonal Cluster Centers. From left to right and top to bottom: Spring, Summer, Fall, and Winter.

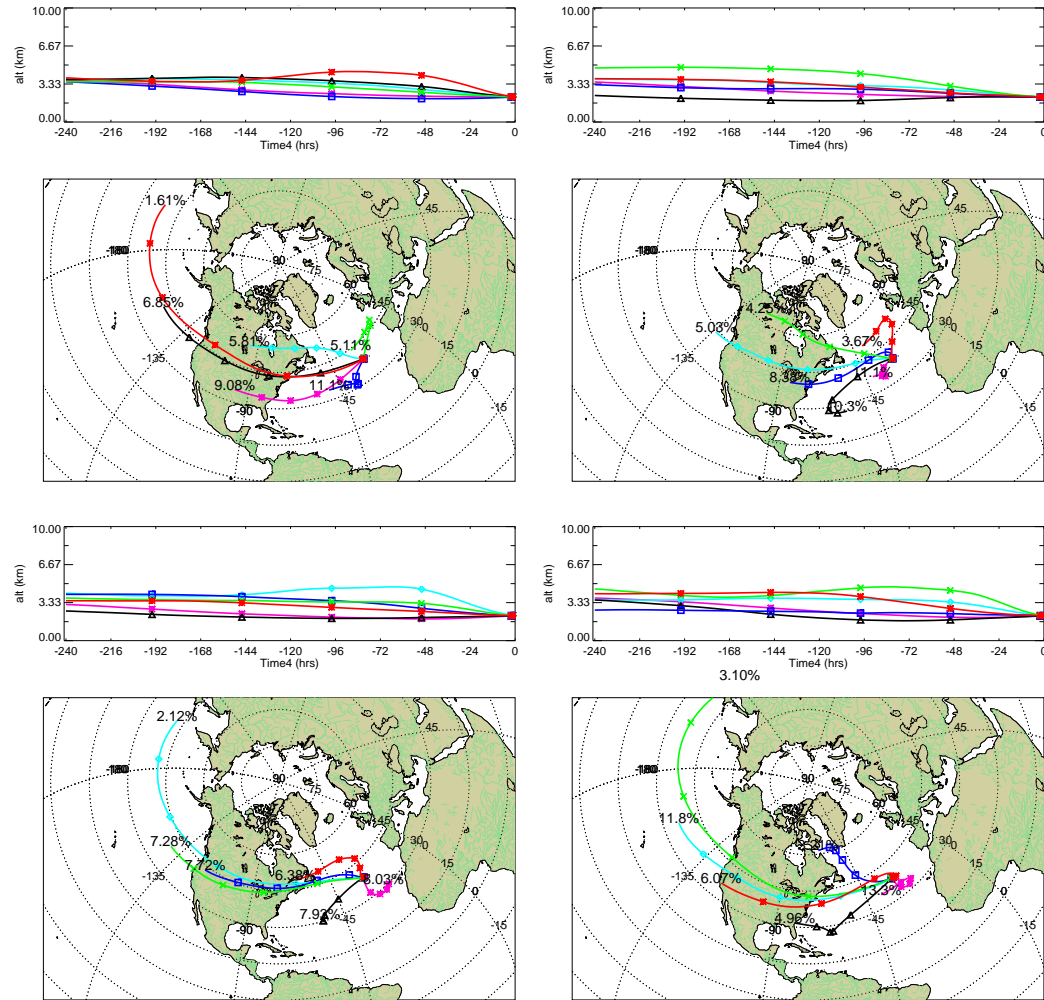


Figure A.8 1960–2006 Non-divergent Seasonal Cluster Centers. From left to right and top to bottom: Spring, Summer, Fall, and Winter.

Table A.2 Variance Comparison for the 1960–2006 Winter Mixed-Flow and Non-divergent Cluster Centers.

Season	Method	Norm. Within Variance	Between Variance	% Diff. NWV	% Diff. BV
Spring	Mixed Flow	24841	222753	7.0	-24.7
	Non-divergent	23094	277689	0.0	0.0
Summer	Mixed Flow	18542	155512	4.0	-13.7
	Non-divergent	17803	176832	0.0	0.0
Fall	Mixed Flow	23323	235312	5.7	-15.5
	Non-divergent	22004	271813	0.0	0.0
Winter	Mixed Flow	27884	290960	5.5	-5.1
	Non-divergent	26428	306453	0.0	0.0

riority, the research goals were also considered. The cluster centers were used for the transport and measurement analyses as described in Chapter 3 for the period of station operation from 2001–2006.

The purpose of the transport analysis was to find the percent of time the station was impacted by the different potential source regions. To accomplish this, all of the trajectories were assigned to the cluster centers for each season of every year, and the fraction of trajectories assigned to each cluster was found. To compare the suitability of each cluster center set for this analysis, the within variance was calculated for each of the cluster center sets with all of the trajectories. We found that the mixed-flow cluster centers had slightly smaller within variance, compared to the non-divergent cluster center within variance when calculated using the entire trajectory set, but the non-divergent within variance was less than 5% greater for all seasons. Although the mixed flow clusters were a better fit to the entire trajectory set, the difference between the two methods was not significant.

To complete the measurement analysis, hourly measurements at the station were assigned to the cluster centers. This was based on the hourly trajectory ensembles,

which indicate the cluster that coincided with the observations at the station at that time. If the trajectory ensemble indicated transport from only one cluster, then the measurement was easily assigned to that cluster. However, if the trajectory ensemble was divergent and indicated two or more clusters, the process of assigning the measurement became complex. The most logical choice for this analysis was the non-divergent method; the ensembles used to assign the measurements were also used to create the cluster centers, and there was higher confidence in assigning the measurements to the cluster which was indicated by all six trajectories arriving at the same time.

Based the variance analysis and the research goals, the best method for this study was the non-divergent technique. The non-divergent cluster centers were the most representative of the member trajectories, were necessary to complete the best measurement analysis, and were nearly equivalent to the mixed-flow cluster centers for the transport analysis.

Appendix B

1960–2006 Spring, Fall, and Winter Frequency Plots at the Pico Mountain Observatory

This appendix contains the spring, fall, and winter frequency plots, which complement the summer frequency plots shown in Figure 3.2. Each frequency plot coincides with one of the seasonal cluster centers, which are shown in Figure 3.1, to reveal the extent of the region from which air might be transported in each cluster.

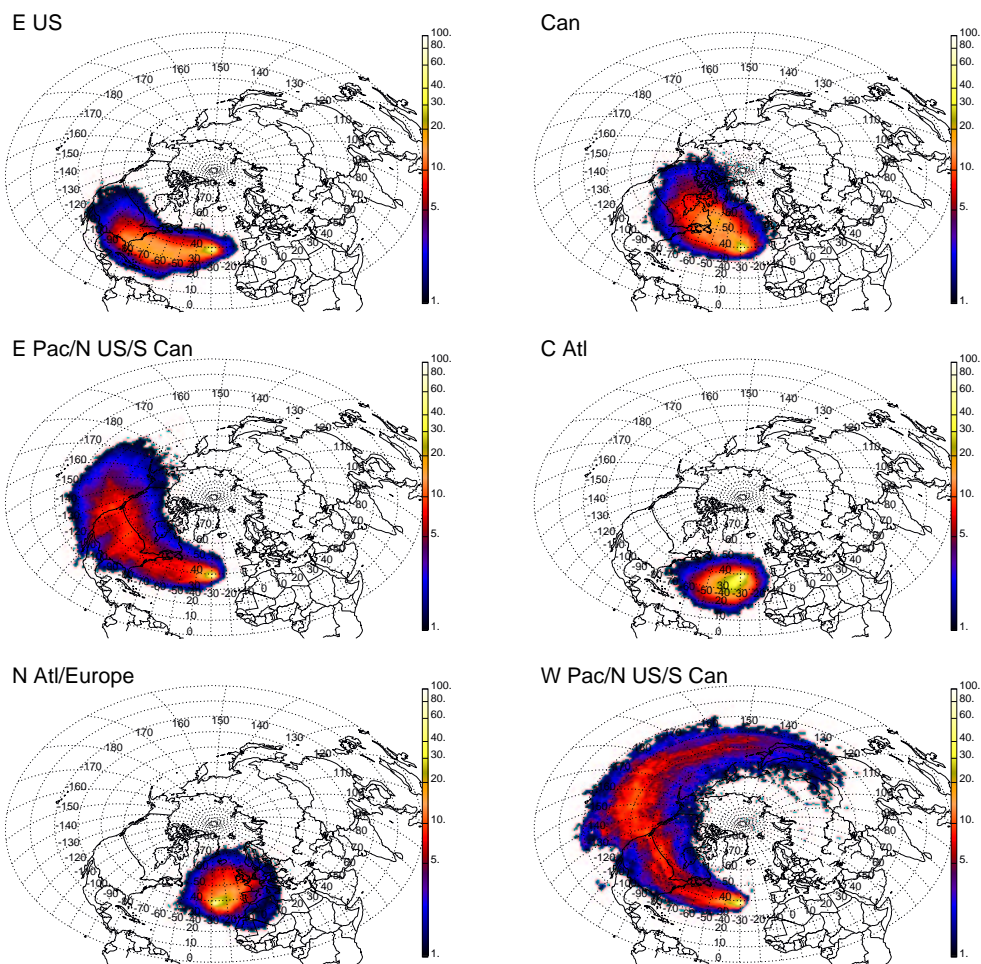


Figure B.1 1960–2006 Spring Cluster Area-Normalized Frequency. Values less than 1% are not shown. The frequency of each grid cell has been multiplied by the ratio of the grid cell area at the equator to the area of the grid cell of interest, so that the frequencies are not biased by latitude.

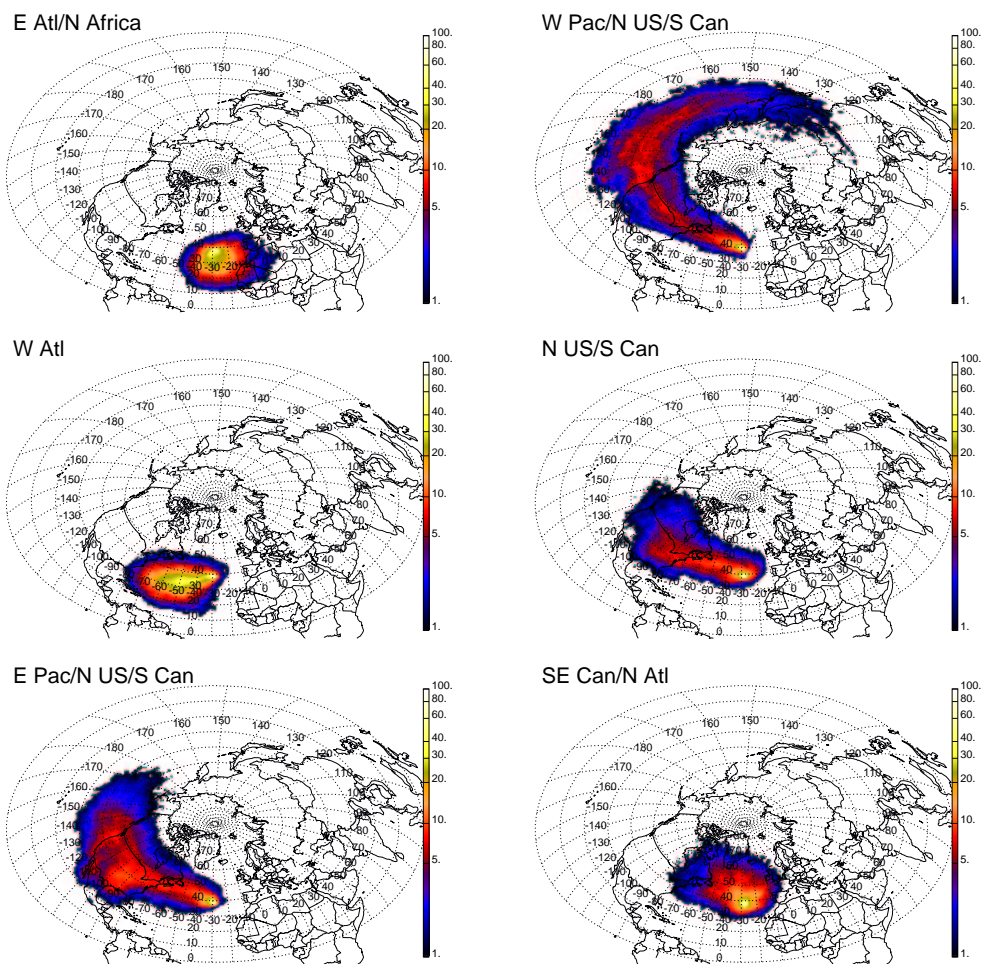


Figure B.2 1960–2006 Fall Cluster Area-Normalized Frequency. Values less than 1% are not shown. The frequency of each grid cell has been multiplied by the ratio of the grid cell area at the equator to the area of the grid cell of interest, so that the frequencies are not biased by latitude.

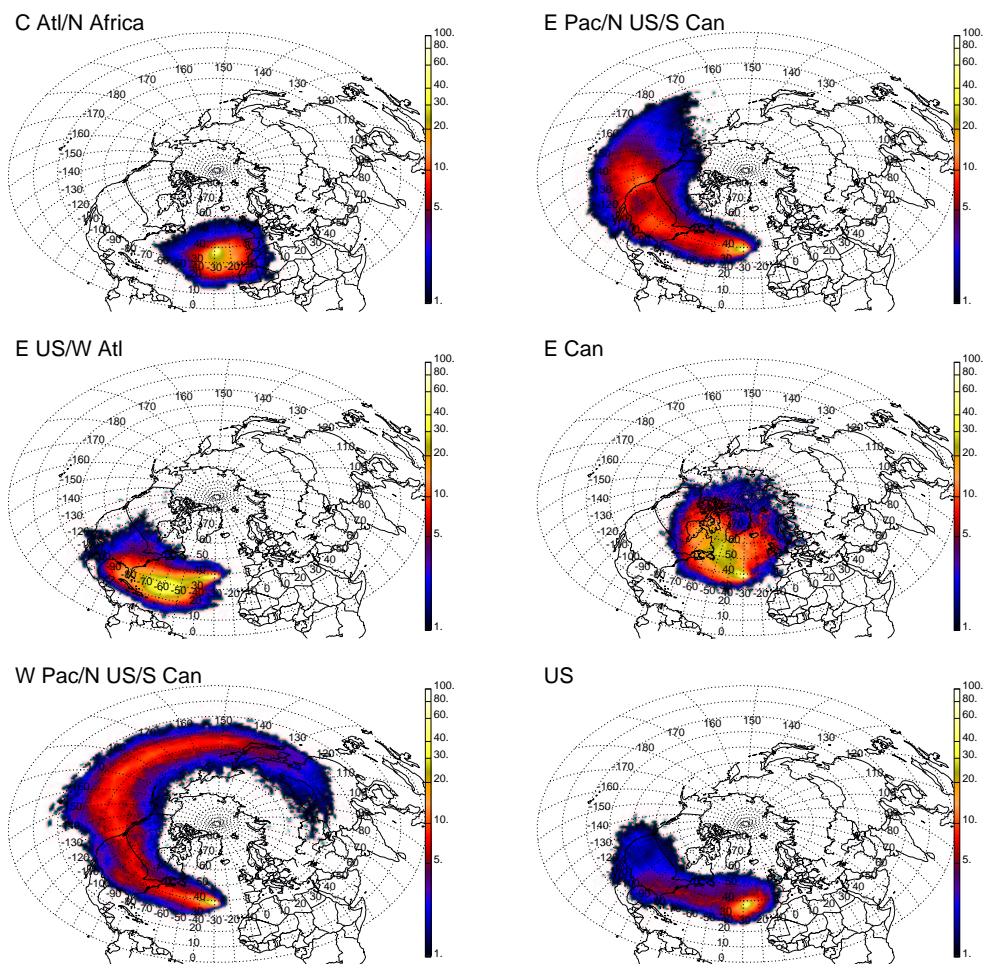


Figure B.3 1960–2006 Winter Cluster Area-Normalized Frequency. Values less than 1% are not shown. The frequency of each grid cell has been multiplied by the ratio of the grid cell area at the equator to the area of the grid cell of interest, so that the frequencies are not biased by latitude.

Appendix C

Spring and Winter Transport Correlations with the NAOI

This appendix contains the plots of the NAOI for spring and winter for 1960–2006 versus the total transport percents for those seasons and years for each cluster center. The correlations presented on these plots were discussed in section 3.2.

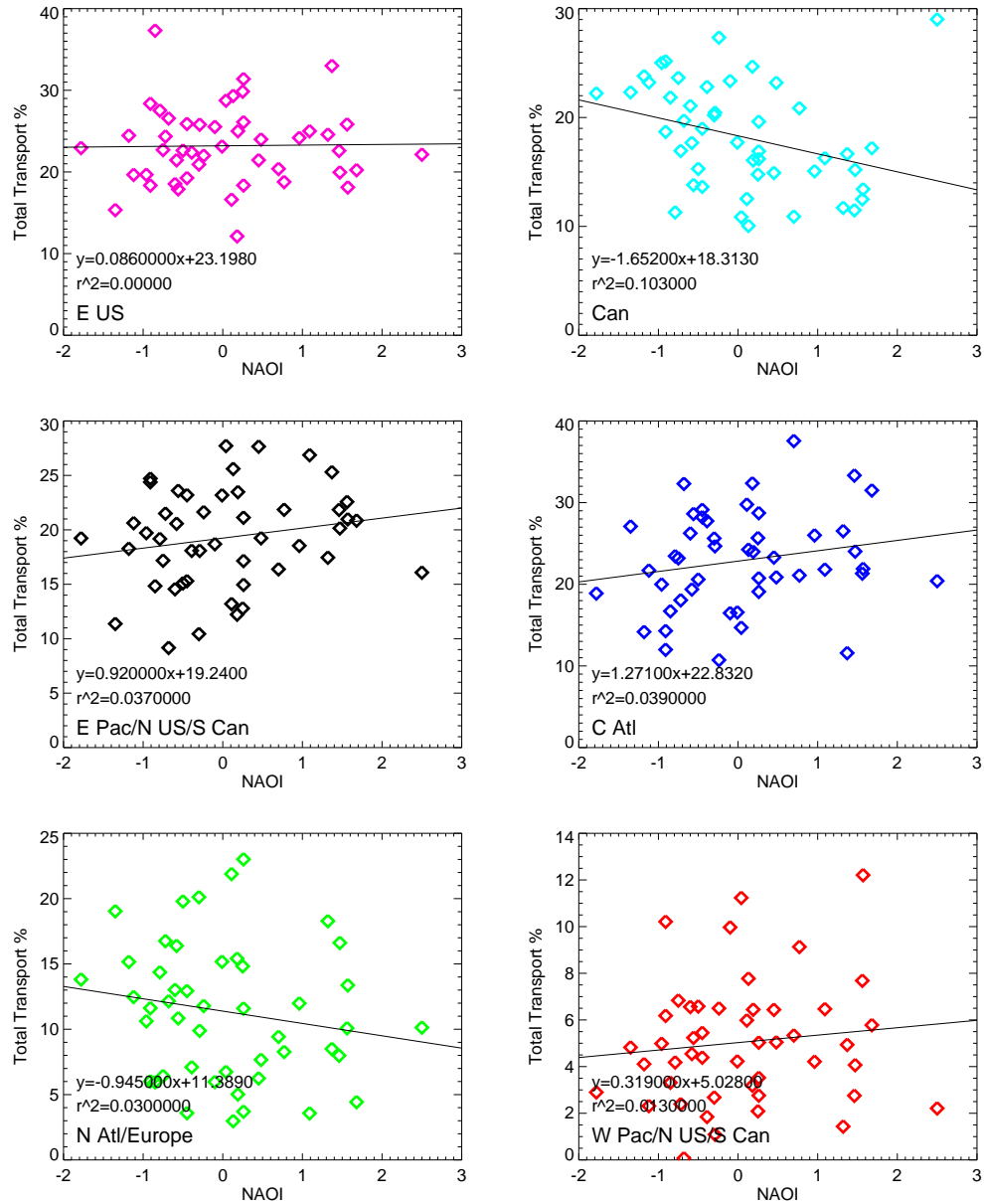


Figure C.1 Linear Regression of the 1960–2006 Spring Cluster Transport and the North Atlantic Oscillation Index (NAOI). Each data point represents the transport and NAOI for an individual year. The best-fit line for the one-sided regression is shown as well as the equation of the line and the squared correlation coefficient.

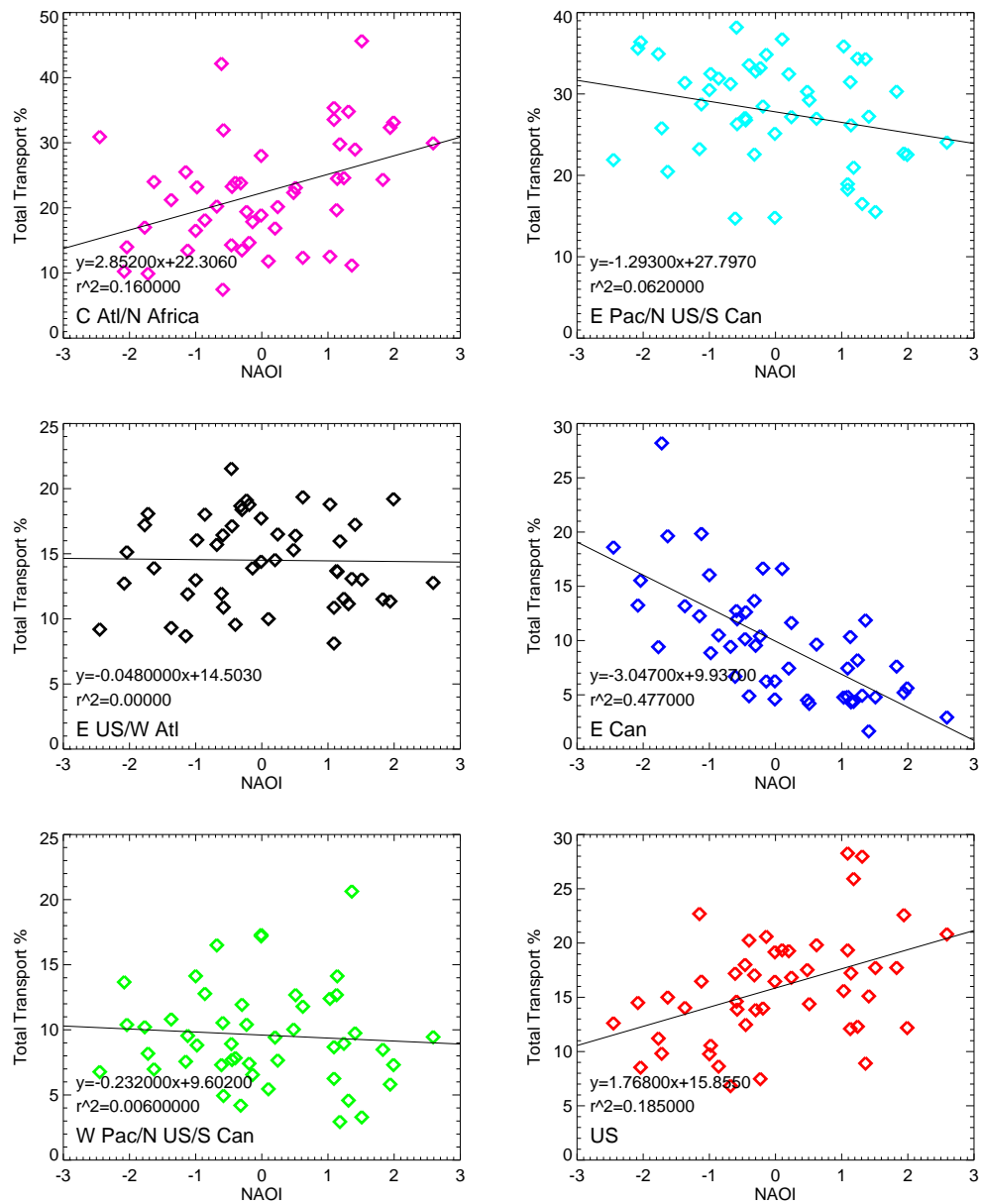


Figure C.2 Linear Regression of the 1960–2006 Winter Cluster Transport and the North Atlantic Oscillation Index (NAOI). Each data point represents the transport and NAOI for an individual year. The best-fit line for the one-sided regression is shown as well as the equation of the line and the squared correlation coefficient.

Appendix D

Time Series of 1960–2006 Seasonal Total Transport Percents

This appendix contains the time series plots of the total transport percents calculated for 1960–2006 for each of the seasonal cluster centers. These plots were created so that any long-term trends in transport could be identified. After analyzing the slopes and correlations of the transport percents, we did not find any strong trends in the transport percentages over time. All seasonal clusters were uncorrelated and had slopes less than 0.1, except for the fall W Atl and winter E Can cluster, which had slope and squared correlation coefficients of -0.145 and 0.134 and -0.138 and 0.124, respectively.

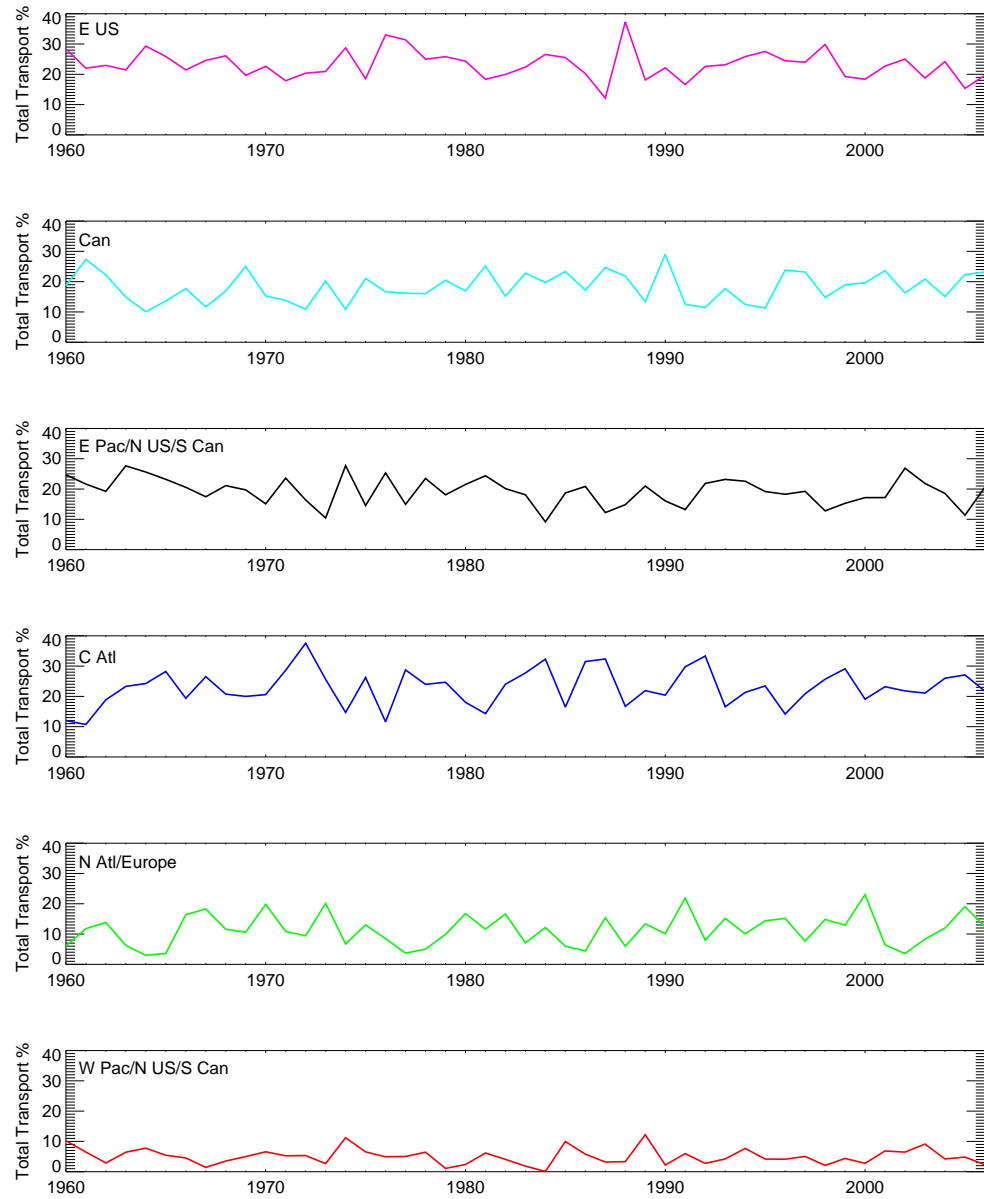


Figure D.1 Time series of the total transport percents for each year in the spring seasonal clusters center from 1960–2006.

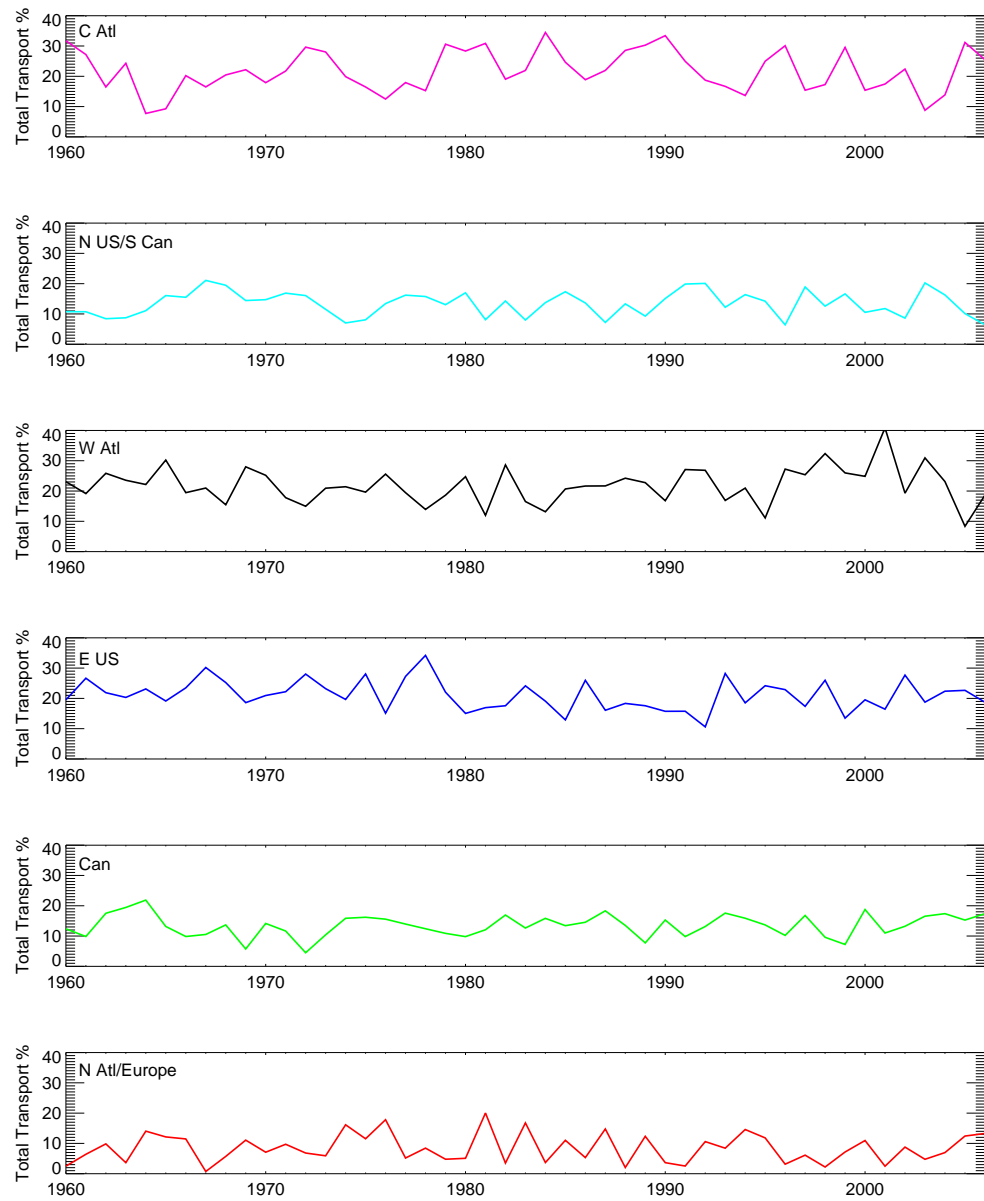


Figure D.2 Time series of the total transport percents for each year in the summer seasonal clusters center from 1960–2006.

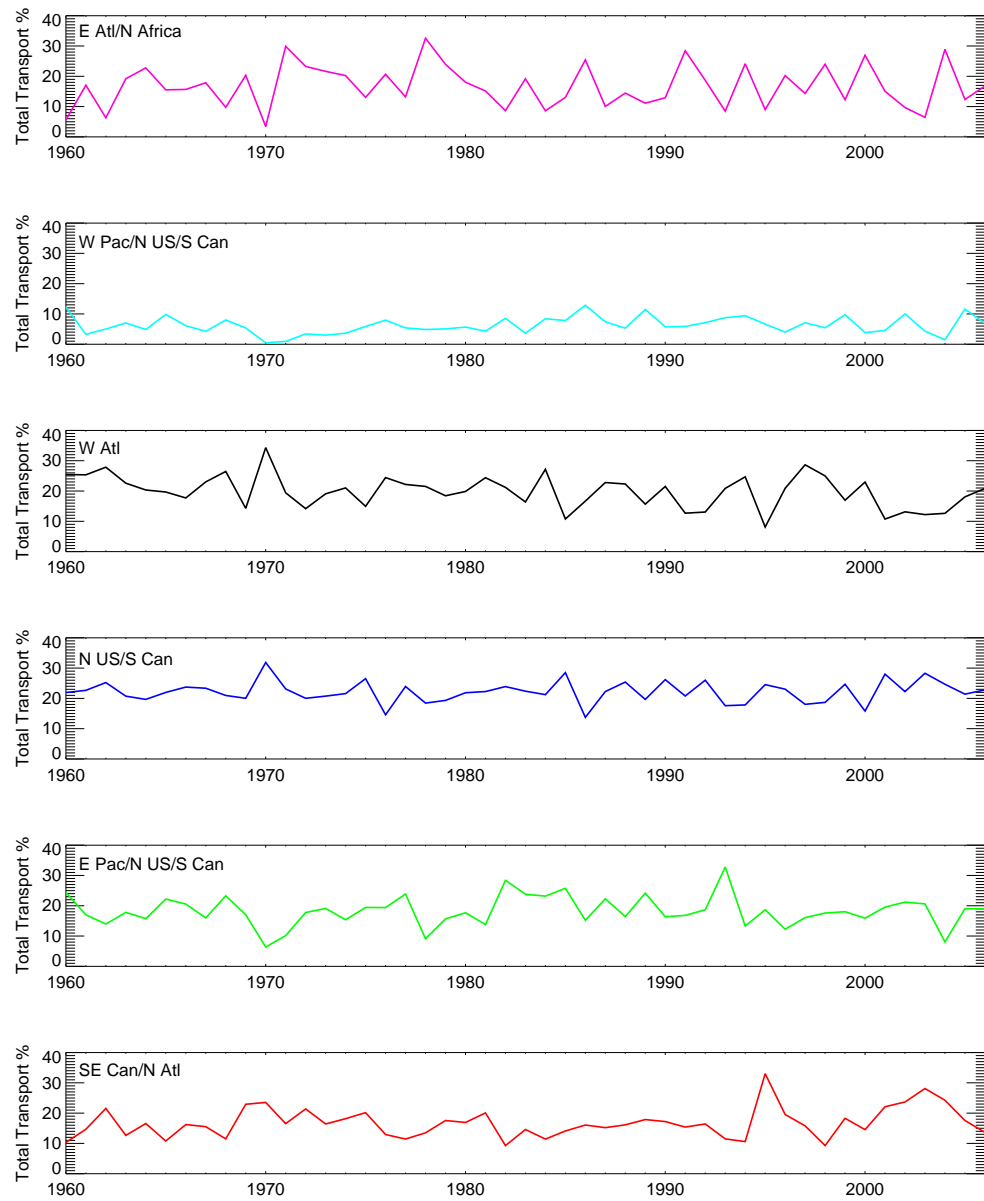


Figure D.3 Time series of the total transport percents for each year in the fall seasonal clusters center from 1960–2006.

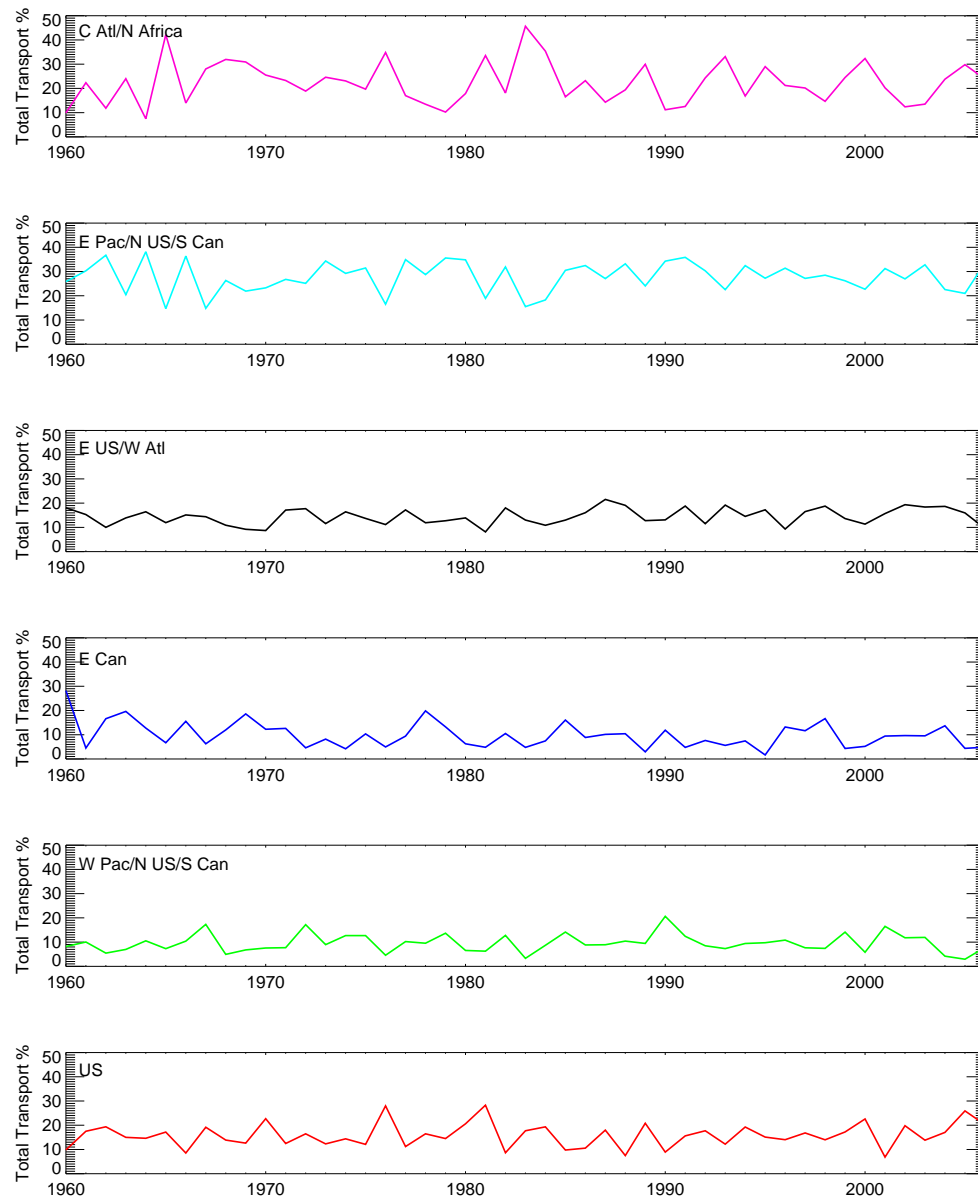


Figure D.4 Time series of the total transport percents for each year in the winter seasonal clusters center from 1960–2006.

Appendix E

Examining the Uncertainty of the 10-Day HYSPLIT Backward Trajectories

Ten-day backward trajectories were used for the clustering analysis. As stated earlier, *Stohl* [1998] found typical trajectory error to be 20% of the distance travelled by a trajectory (error increases with increasing trajectory length). At the end of a 10-day trajectory, the trajectory’s location could be more than 1,000 kilometers away from the “true” position of the air parcel that was tracked. In the majority of recent published work that employed a trajectory clustering technique, the trajectory lengths utilized were five days or less, so that such large potential errors were avoided. To identify distant upwind source regions for the Pico Mountain observatory, however, the use of 10-day backward trajectories was necessary as shorter backward trajectories do not travel very far out of the Atlantic region, particularly in summer.

A few steps were already taken to reduce the effects of trajectory uncertainty on these results. The clustering analysis and the large trajectory data set used (1960–2006) helped minimize the error of the individual trajectories through the averaging

process used to generate the cluster centers. The non-divergent clustering technique prevented mixed-flow ensembles, which contain the trajectories most likely affected by error, from being used to calculate the final cluster centers. Additionally, the transport regions identified by the clustering analysis were considered as the areas defined by the frequency plots (Figures 3.2 and B.1-B.3), which were thousands of square kilometers in size (on a similar scale to the anticipated error in the individual trajectories). Although those steps compensated for the error of the 10-day backward trajectories, we also wanted to verify that the final cluster center locations were not more heavily dependent on the first five days of transport, due to the greater potential error, than the last five days of transport (five days directly preceding the station arrival time).

To test this, we assigned the 2001–2006 hourly trajectories to the 1960–2006 seasonal cluster centers, following the method used in section 3.2, but instead of assigning the trajectories to the nearest cluster center based on the distance summed over all 10 days of transport, the trajectories were assigned to the nearest cluster center based on distance summed over only the last five days of transport. Total transport, which includes mixed-flow times, and non-divergent transport, which does not include mixed-flow times, were calculated in this way. The 5-day total and non-divergent transport percents were then regressed against the 10-day total and non-divergent transport percents on a seasonal basis. If both the first and last five days of transport were equally important in generating the final cluster centers, we expected that the 5 and 10-day percents would be correlated. Figures E.1 and E.2 show the regressions.

We found that the 5-day and 10-day transport percents were correlated ($r^2 = 0.41 - 0.85$). This indicated that the potentially greater error in the first five days of transport of the 10-day trajectories did cause the final cluster centers to deviate from the path indicated by the last five days of transport, but only to a moderate degree. If the error in the first five days of transport had caused the cluster centers to shift, the 5-day and 10-day transport percent would have been uncorrelated. Therefore the use

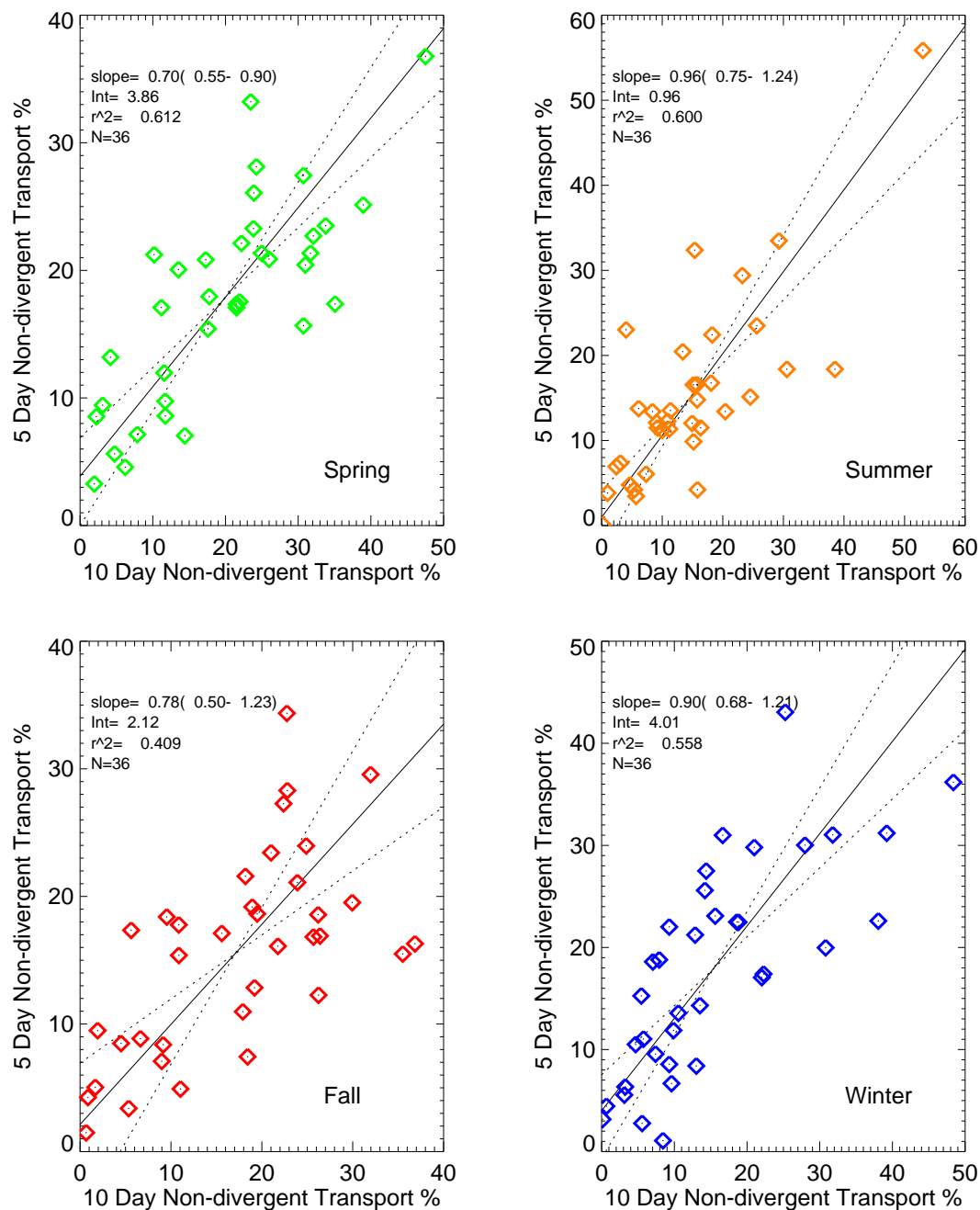


Figure E.1 Linear Regression of 5-Day Non-divergent Transport Percents against 10-Day Non-divergent Transport Percents. The two-sided regressions of the transport percents for all the years and clusters are performed on a seasonal basis; the slope of the best-fit line and the correlation coefficient squared are shown in the upper left hand corner of the plots. The dashed lines show the ordinary least-squares regression of x versus y and y versus x. For each year, the non-divergent percents were calculated by dividing the number of non-divergent trajectories assigned to each cluster by the total number of non-divergent trajectories for the season.

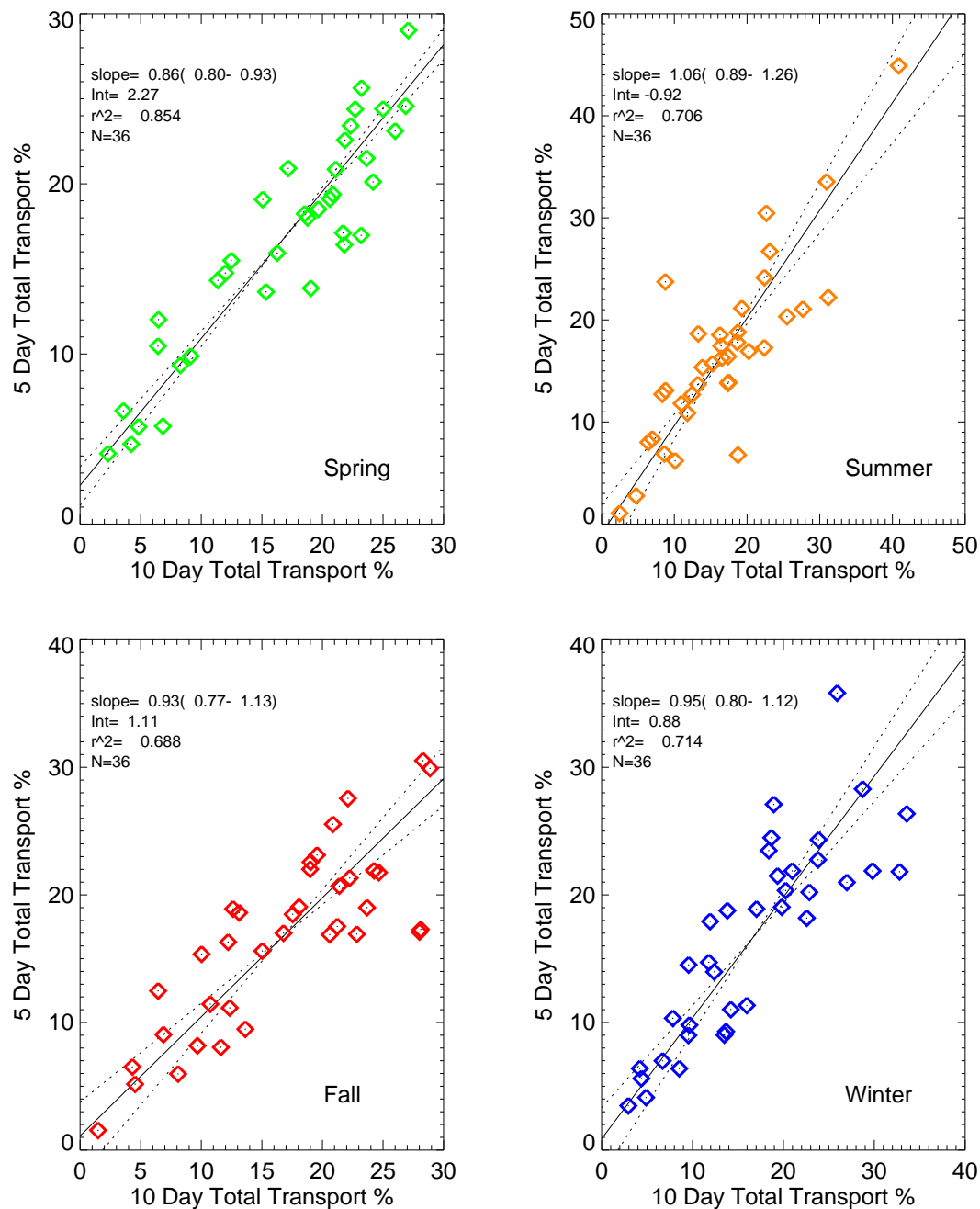


Figure E.2 Linear Regression of 5-Day Total Transport Percents against 10-Day Total Transport Percents. The two-sided regressions of the transport percents for all the years and clusters are performed on a seasonal basis; the slope of the best-fit line and the correlation coefficient squared are shown in the upper left hand corner of the plots. The dashed lines show the ordinary least-squares regression of x versus y and y versus x.

of 10-day trajectories in the clustering analysis produced meaningful cluster centers, even though the possible error in the long-distance trajectories can be great.

Appendix F

Analysis of Upslope-Flow Screening on the CO and O₃ Measurements

This appendix contains plots displaying the CO and O₃ means presented in section 3.3 and the CO and O₃ means which were screened for upslope flow periods, based on *Kleissl et al.* [2007]. These plots were discussed in section 2.3. The observed changes in the ranking of the cluster centers by mixing ratios occur for 2004 spring O₃, 2005 summer CO, and 2005 winter CO. In spring of 2004, the screened O₃ mean for the E Pac/N US/S Can cluster is not significantly different from the N Atl/Europe. The highest spring 2004 O₃ then coincided with transport from the N Atl/Europe, in addition to the E Pac/N US/S Can. In summer of 2005, the screened CO means associated with transport from the N Atl/Europe and Can decrease slightly to overlap the increased E US CO mean; therefore, the highest CO means coincide with transport from the E US, in addition to the previously observed clusters. The screened 2005 winter CO means are not significantly different between cluster centers; this follows the previous conclusion that winter is characterized by low variability.

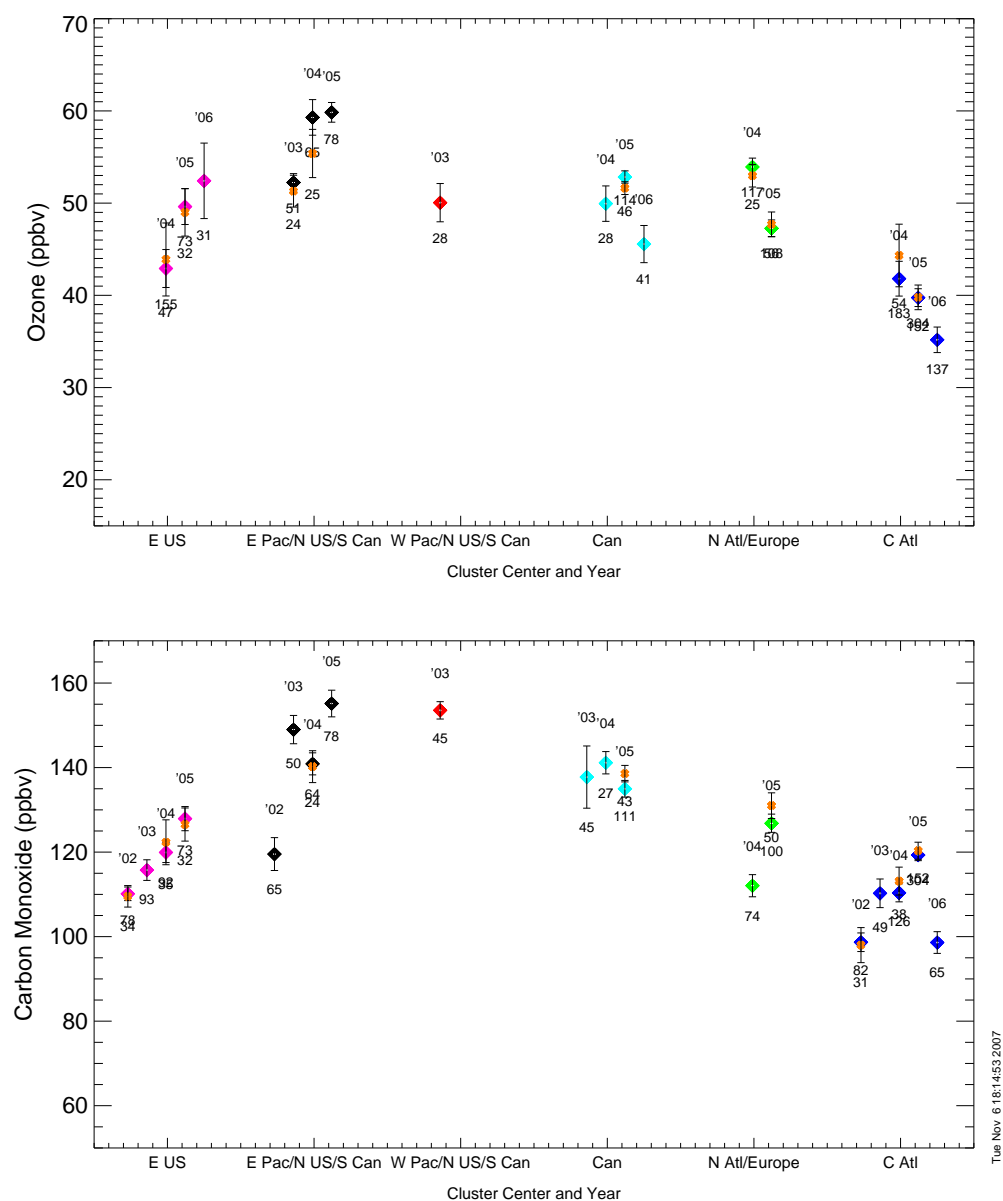


Figure F.1 Comparison of upslope flow screened and unscreened spring CO and O₃ means by cluster center. The unscreened means, which were used for the analyses presented in this thesis are shown in their original colors. The upslope flow screened means are shown in orange. The coinciding year and number of points used to calculate each mean are shown above and below the means, respectively.

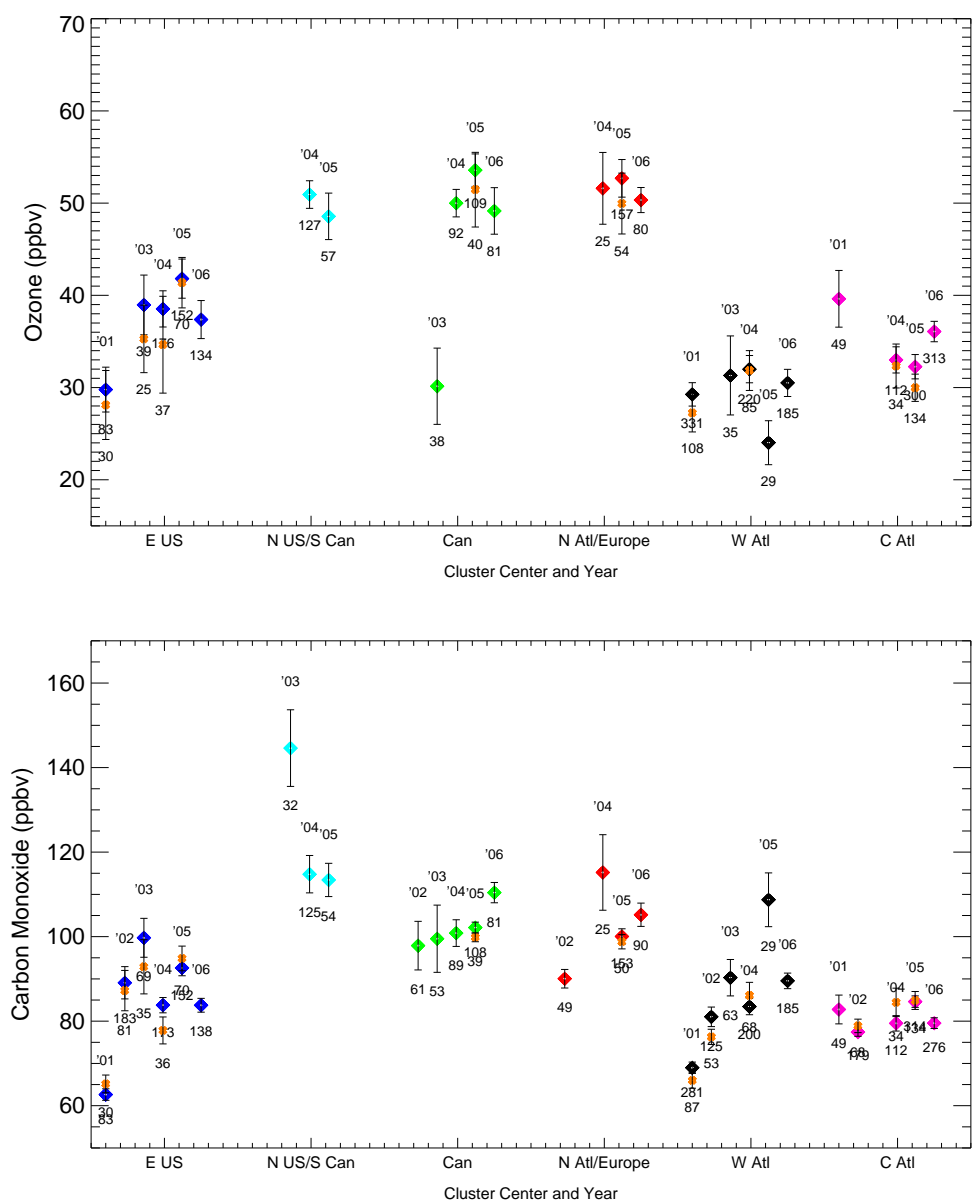


Figure F.2 Comparison of upslope flow screened and unscreened summer CO and O₃ means by cluster center. The unscreened means, which were used for the analyses presented in this thesis are shown in their original colors. The upslope flow screened means are shown in orange. The coinciding year and number of points used to calculate each mean are shown above and below the means, respectively.

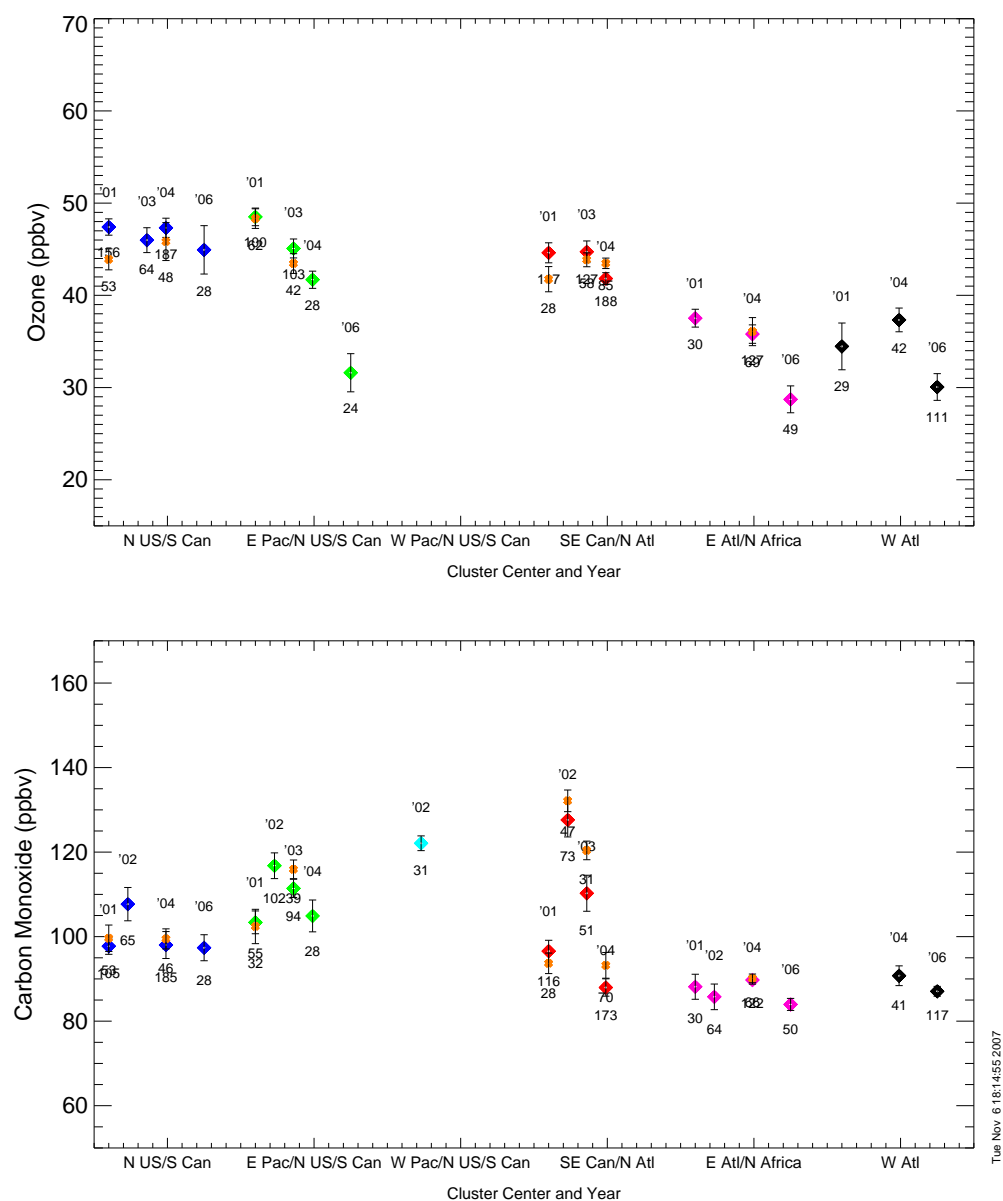


Figure F.3 Comparison of upslope flow screened and unscreened fall CO and O₃ means by cluster center. The unscreened means, which were used for the analyses presented in this thesis are shown in their original colors. The upslope flow screened means are shown in orange. The coinciding year and number of points used to calculate each mean are shown above and below the means, respectively.

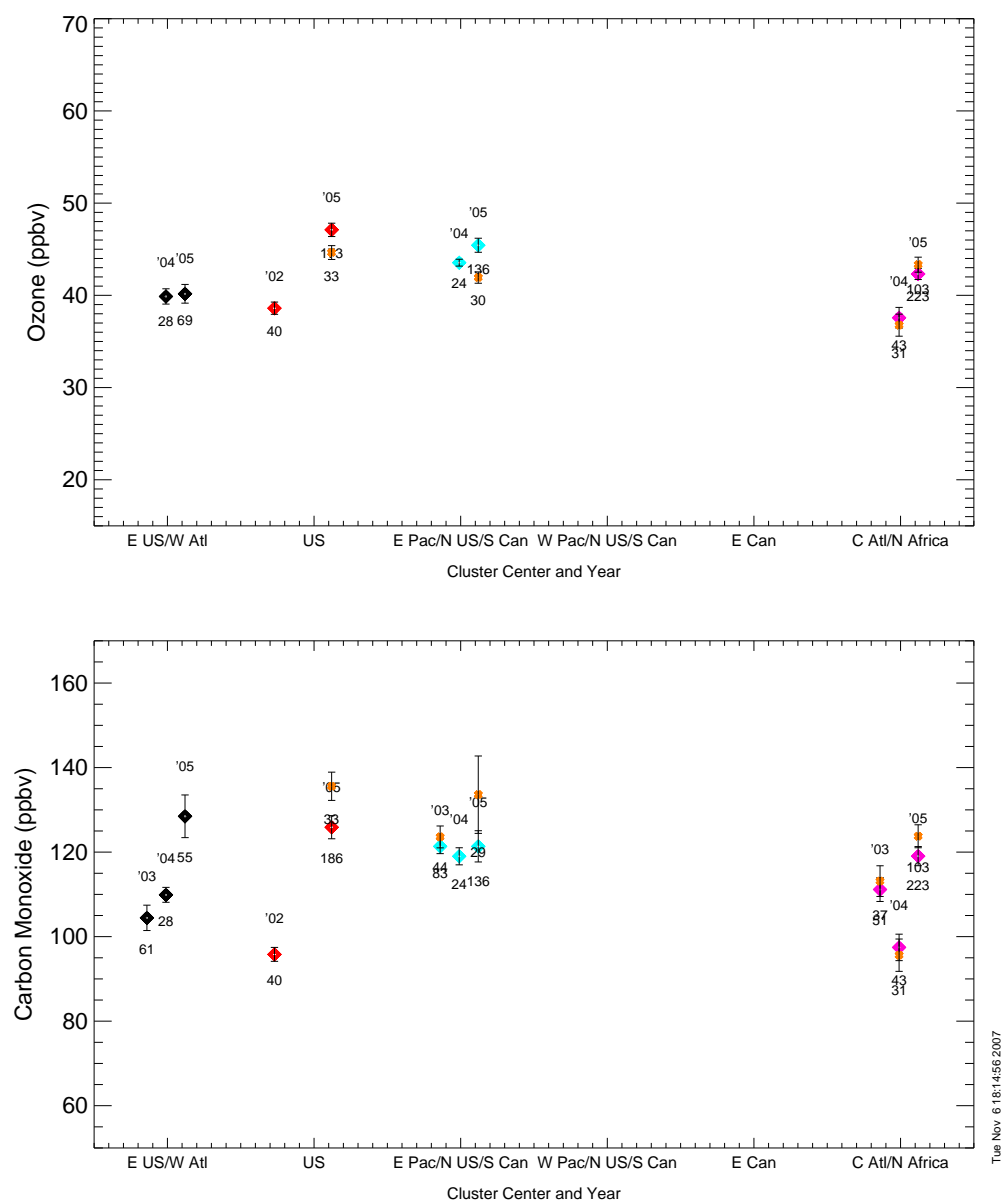


Figure F.4 Comparison of upslope flow screened and unscreened winter CO and O₃ means by cluster center. The unscreened means, which were used for the analyses presented in this thesis are shown in their original colors. The upslope flow screened means are shown in orange. The coinciding year and number of points used to calculate each mean are shown above and below the means, respectively.

References

Abdalmogith, S., and R. Harrison, The use of trajectory cluster analysis to examine the long-range transport of secondary inorganic aerosol in the UK, *Atmospheric Environment*, *39*, 6686–6695, 2005.

Bey, I., D. Jacob, J. Logan, and R. Yantosca, Asian chemical outflow to the Pacific in spring: Origins, pathways, and budgets, *Journal of Geophysical Research*, *106*, 23,097–23,113, 2001.

Borge, R., J. Lumbreras, S. Vardoulakis, P. Kassomenos, and E. Rodríguez, Analysis of long-range transport influences on urban PM₁₀ using two-stage atmospheric trajectory clusters, *Atmospheric Environment*, *41*, 4434–4450, 2007.

Cape, J., J. Methven, and L. Hudson, The use of trajectory cluster analysis to interpret trace measurement at Mace Head, Ireland, *Atmospheric Environment*, *34*, 3651–3663, 2000.

CIFFC, 2001-2006 Canada Report on the wildland fire and fire response activity, Report access via Canadian Interagency Forest Fire Centre Website (<http://www.ciffc.ca/web/hotpoints/reports.shtml>), Date Accessed: July 2007, 2007.

Draxler, R. R., and G. Hess, An overview of the HYSPLIT 4 modeling system for trajectories, dispersion, and deposition, *Australian Meteorological Magazine*, *47*, 295–308, 1998.

Draxler, R. R., and G. D. Rolph, HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory), Model access via NOAA ARL READY Website (<http://www.arl.noaa.gov/ready/hysplit4.html>), NOAA, Air Resources Laboratory, Silver Spring, MD., Date Accessed: April 2007, 2003.

ECMWF, Users guide to ECMWF products 4.0, *Tech. Rep. Meteorol. Bull. M3.2*, European Centre for Medium-Range Weather Forecasts, Reading, U.K., 2005.

Eneroth, K., E. Kjellström, and K. Holmén, A trajectory climatology for Svalbard; investigating how atmospheric flow patterns influence observed tracer concentrations, *Physics and Chemistry of the Earth*, 28, 1191–1203, 2003.

Finlayson-Pitts, B., and J. Pitts, *Atmospheric Chemistry: Fundamentals and Experimental Techniques*, John Wiley and Sons, Inc., Canada, 1986.

Harris, J., and J. Kahl, A descriptive atmospheric transport climatology for the Mauna Loa Observatory, using clustered trajectories, *Journal of Geophysical Research*, 95, 13,651–13,667, 1990.

Hartigan, J. A., *Clustering algorithms*, John Wiley and Sons, Inc., New York, NY, 1975.

Honrath, R. E., et al., Regional and hemispheric impacts of anthropogenic and biomass burning emissions on summertime CO and O₃ in the North Atlantic lower free troposphere, *Journal of Geophysical Research*, 109, doi: 10.1029/2004JD005147, 2004.

Hurrell, J., NAO Index Data provided by the Climate Analysis Section (<http://www.cgd.ucar.edu/cas/jhurrell/indices.data.html>), NCAR, Boulder, USA, Date Accessed: August 2007, 1995.

Hurrell, J., North Atlantic Oscillation (NAO), *Encyclopedia of Ocean Sciences*, 4, 1904–1911, doi:10.1006/rwos.2001.0263, 2001.

Jaeglé, L., et al., Sources and budgets for CO and O₃ in the northeastern Pacific during the spring of 2001: Results from the PHOBEA-II Experiment, *Journal of Geophysical Research*, *108*, doi: 10.1029/2002JD003121, 2003.

Jaffe, D., et al., Long-range transport of Siberian biomass burning emissions and impact on surface ozone in western North America, *Geophysical Research Letters*, *31*, doi: 10.1029/2004GL020093, 2004.

Jorba, O., C. Pérez, F. Rocadenbosch, and J. Baldasano, Cluster Analysis of 4-Day Back Trajectories Arriving in the Barcelona Area, Spain from 1997 to 2002, *Journal of Applied Meteorology*, *43*, 887–901, 2004.

Kasischke, E., et al., Influences of boreal fire emissions on Northern Hemisphere atmospheric carbon and carbon monoxide, *Global Biogeochemical Cycles*, *19*, doi:10.1029/GB002300, 2005.

Kleissl, J., R. Honrath, M. P. Dziobak, D. Tanner, M. Val Martín, R. Owen, and D. Helmig, Occurrence of upslope flows at the Pico mountaintop observatory: A case study of orographic flows on a small, volcanic island, *Journal of Geophysical Research*, *112*, doi:10.1029/2006JD007565, 2007.

Lapina, K., R. Honrath, R. Owen, M. Val Martín, and G. Pfister, Evidence of significant large-scale impacts of boreal fires on ozone levels in the midlatitude Northern Hemisphere free troposphere, *Geophysical Research Letters*, *33*, doi:10.1029/2006GL025878, 2006.

Lapina, K., R. Honrath, R. Owen, M. Val Martín, E. Hyer, P. Fialho, and F. Barata, Evidence of a late-summer decrease in /nox CO emission ratios from boreal fires, *Journal of Geophysical Research*, submitted to JGR, 2007.

Li, F., A. Vogelmann, and V. Ramanathan, Saharan Dust Aerosol Radiative Forcing Measured from Space, *Journal of Climate*, *17*, 2558–2571, 2004.

NCEP/NCAR, Archived Reanalysis Meteorological Data, Data access via NOAA ARL archived meteorological data website (www.arl.noaa.gov/pub/archives/reanalysis), NOAA, Air Resources Laboratory, Silver Spring, MD., Date Accessed: April 2007, 2007.

NICC, Incident Management Reports, Report access via National Interagency Coordination Center Website (<http://iys.cidi.org/wildfire/index.html>), Date Accessed: July 2007, 2007.

Niemi, J., H. Tervahattu, A. Virkkula, R. Hillamo, K. Teinilä, K. Koponen, and M. Kulmala, Continental impact on the marine boundary layer coarse particles over the Atlantic Ocean between Europe and Antarctica, *Atmospheric Research*, *75*, 301–321, doi:10.1016/j.atmosres.2005.01.005, 2005.

NIFC, Wildland Fire Season 2005, Information access via National Interagency Fire Center Website (http://www.nifc.gov/fire_info/fire_summaries/summary_2005.htm), Date Accessed: July 2007, 2007.

NOAA/NWS, Climate Prediction Center: Northern Hemispheric Teleconnection Patterns, Information access via National Weather Service Website (<http://www.cpc.ncep.noaa.gov/data/teledoc/telecontents.shtml>), Date Accessed: July 2007, 2005.

Olivier, J., and J. Berdowski, Global emissions sources and sinks, in *The Climate System*, edited by J. Berdowski, R. Guicherit, and B. Heij, pp. 33–78, A.A. Balkema, Brookfield, Vt., 2001.

Owen, R., O. Cooper, A. Stohl, and R. Honrath, An analysis of the mechanisms of North American pollutant transport to the central North Atlantic lower free

troposphere, *Journal of Geophysical Research*, 111, doi: 10.1029/2006JD007062, 2006.

Owen, R. C., A Climatological Study of Transport to the PICO-NARE Site Using Atmospheric Backward Trajectories, Master's thesis, Michigan Technological University, 2003.

Sánchez-Ccoyllo, O., P. L. Silva Dias, M. de Fátima Andrade, and S. Freitas, Determination of O_3 -, CO- and PM_{10} -transport in the metropolitan area of São Paulo, Brazil through synoptic-scale analysis of back trajectories, *Meteorology and Atmospheric Physics*, 92, 83–93, doi: 10.1007/s00703-005-0139-6, 2006.

Seibert, P., and A. Frank, Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, *Atmospheric Chemistry and Physics*, 4, 51–63, 2004.

Seinfeld, J., and S. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, second ed., John Wiley and Sons, Inc., 2006.

Soja, A. J., et al., Climate-induced boreal forest change: Predictions versus current observations, *Global and Planetary Change*, 56, 274–296, 2007.

Stohl, A., Computation, Accuracy and Applications of Trajectories - A Review and Bibliography, *Atmospheric Environment*, 32, 947–966, 1998.

Stohl, A., C. Forster, A. Frank, P. Seibert, and G. Wotawa, Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, *Atmospheric Chemistry and Physics*, 5, 2461–2474, 2005.

Val Martín M., R. Honrath, R. Owen, G. Pfister, P. Fialho, and F. Barata, Significant enhancements of nitrogen oxides, black carbon and ozone in the North Atlantic lower free troposphere resulting from North American boreal wildfires, *Journal of Geophysical Research*, 111, doi: 10.1029/2006JD007530, 2006.

Val Martín M., R. Honrath, R. Owen, K. Lapina, and Q. Li, Seasonal variation of nitrogen oxides in the central North Atlantic lower free troposphere, *Journal of Geophysical Research*, Manuscript in Preparation, 2007.

Wang, Y., X. Zhang, and R. Arimoto, The contribution from distant dust sources to the atmospheric particulate matter loadings at XiAn, China during spring, *Science of the Total Environment*, 368, 875–883, doi: 10.1016/j.scitotenv.2006.03.040, 2006.